FIRST PASSAGE PERCOLATION ON THE NEWMAN-WATTS SMALL WORLD MODEL

JÚLIA KOMJÁTHY[1,*] AND VIKTÓRIA VADON[2]

ABSTRACT. The Newman-Watts model is given by taking a cycle graph of n vertices and then adding each possible edge $(i,j), |i-j| \neq 1 \mod n$ with probability ρ/n for some $\rho > 0$ constant. In this paper we add i.i.d. exponential edge weights to this graph, and investigate typical distances in the corresponding random metric space given by the least weight paths between vertices. We show that typical distances grow as $\frac{1}{\lambda} \log n$ for a $\lambda > 0$ and determine the distribution of smaller order terms in terms of limits of branching process random variables. We prove that the number of edges along the shortest weight path follows a Central Limit Theorem, and show that in a corresponding epidemic spread model the fraction of infected vertices follows a deterministic curve with a random shift.

1. THE MODEL AND MAIN RESULTS

1.1. The Newman-Watts model.

The Newman-Watts small world model, often referred to as "small world" in short, is one of the first random graph models created to model real-life networks. It was introduced by Watts and Strogatz [34], and a simplifying modification was made by Newman and Watts [29] later. The Newman-Watts model consist of a cycle on n vertices, each connected to the $k \geq 1$ nearest vertices, and then extra shortcut edges are added in a similar fashion to the creation of the Erdős-Rényi graph [20]: i.e., for each pair of not yet connected vertices, we connect them independently with probability p.

The model has been studied from different aspects. Newman *et al.* studied distances [30, 31] with simulations and mean-field approximation, as well as the threshold for a large outbreak of the spread of non-deterministic epidemics [28]. Barbour and Reinert treated typical distances rigorously. First, in [6], they studied a continuous circle with circumference n instead of a cycle on n many vertices, and added $\text{Poi}(n\rho/2)$ many 0-length shortcuts at locations chosen according to the uniform measure on the circle. Then, in [7], they studied the discrete model, with all edge lengths equal to 1. They showed that typical distances in both models scale as $\log n$.

Besides typical distances, the mixing time of simple random walk on the Newman-Watts model was also studied, i.e., the time when the distribution of the position of the walker gets close enough to the stationary distribution in total variation distance. Durrett [19] showed that the order of the mixing time is between $(\log n)^2$ and $(\log n)^3$, then Addario-Berry and Lei [1] proved that Durett's lower bound is sharp.

1 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY
2 DEPARTMENT OF STOCHASTICS, BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
E-mail addresses: j.komjathy@tue.nl, vadonvik@math.bme.hu.

1.2. **Main results.** We work on the Newman-Watts small world model [29] with independent random edge weights: we take a cycle C_n on n vertices, that we denote by $[n] := \{1, 2, \ldots, n\}$, and each edge $(i,j) \in [n], |i - j| = 1 \mod n$ is present. Then independently for each $i, j \in [n], |i - j| \neq 1 \mod n$ we add the edge (i,j) with probability ρ/n to form shortcut edges. The parameter ρ is the asymptotic average number of shortcuts from a vertex. Conditioned on the edges of the resulting graph, we assign weights that are i.i.d. exponential random variables with mean 1 to the edges. We denote the weight of edge e by X_e. We write $\mathrm{NW}_n(\rho)$ for a realization of this weighted random graph.

We define the distance between two vertices in $\mathrm{NW}_n(\rho)$ as the sum of weights along the shortest weight path connecting the two vertices. In this respect, the weighted graph with this distance function is a (non-Euclidean) random metric space. Further, interpreting the edge weights as time or cost, the distance between two vertices can also correspond to the time it takes for information to spread from one vertex to the other on the network, or it can model the cost of transmission between the two vertices.

We say that a sequence of events $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ happens with high probability (w.h.p.) if $\lim_{n \to \infty} \mathbb{P}(\mathcal{E}_n) = 1$, that is, the probability that the event holds tends to 1 as the size of the graph tends to infinity. We write $\mathrm{Bin}, \mathrm{Poi}, \mathrm{Exp}$ for binomial, Poisson, and exponential distributions. For random variables $\{X_n\}_{n \in \mathbb{N}}, X$, we write $X_n \xrightarrow{d} X$ if X_n tends to X in distribution as $n \to \infty$. The *moment generating function* of a random variable X is the function $M_X(\vartheta) := \mathbb{E}[\exp\{-\vartheta X\}]$.

Our first result is about typical distances in the weighted graph. Let Γ_{ij} denote the set of all paths γ in $\mathrm{NW}_n(\rho)$ between two vertices $i, j \in [n]$. Then the weight of the shortest weight path is defined by

$$\mathcal{P}_n(i,j) := \min_{\gamma \in \Gamma_{ij}} \sum_{e \in \gamma} X_e. \tag{1.1}$$

Theorem 1.1 (Typical distances). *Let U, V be two uniformly chosen vertices in $[n]$. Then, the distance $\mathcal{P}_n(U, V)$ in $\mathrm{NW}_n(\rho)$ with i.i.d. $\mathrm{Exp}(1)$ edge weights satisfies w.h.p.*

$$\mathcal{P}_n(U, V) - \frac{1}{\lambda} \log n \xrightarrow{d} -\frac{1}{\lambda}(\log W^U W^V + \Lambda + c),$$

where λ is the largest root of the polynomial $p(x) = x^2 + (1 - \rho)x - 2\rho$, Λ is a standard Gumbel random variable, the random variables W^U, W^V are independent copies of the martingale limit of the multi-type branching process defined below in Section 2.3, and $c := \log(1 - \pi_R^2/2) - \log(\lambda(\lambda + 1))$ with $\pi_R = 2/(\lambda + 2)$.

Let us write $\gamma^\star = \gamma^\star(i, j)$ for the path that minimizes the weighted distance in (1.1). We call $\mathcal{H}_n(U, V) := |\gamma^\star(U, V)|$ the *hopcount*, i.e., the number of edges along the shortest-weight path between two uniformly chosen vertices.

Theorem 1.2 (Central Limit Theorem for the hopcount). *Let U, V be two uniformly chosen vertices in $[n]$. Then, the hopcount $\mathcal{H}_n(U, V)$ in $\mathrm{NW}_n(\rho)$ with i.i.d. $\mathrm{Exp}(1)$ edge weights satisfies w.h.p.*

$$\frac{\mathcal{H}_n(U, V) - \frac{\lambda+1}{\lambda} \log n}{\sqrt{\frac{\lambda+1}{\lambda} \log n}} \xrightarrow{d} Z,$$

where Z is a standard normal random variable.

Our next result characterises the proportion of vertices within distance t away from a uniformly chosen vertex U as a function of t. To put this result into perspective, note that we can model the spread of information starting from some *source set* $I_0 \subset [n]$ at time

$t = 0$ as follows: We assume that once a vertex v receives the information at time t, it starts transmitting the information towards all its neighbors at rate 1. Let us denote the vertices that are connected to v by an edge by $H(v)$, then, for each $w \in H(v)$, w receives the information from v at time $t + X_{(v,w)}$. We further assume that transmission happens only after the first receipt of the information, that is, any consecutive receipts are ignored. If instead of the spread of information spread, we model the spread of a disease, this model is often called an SI-epidemic (susceptible-infected).

In the next theorem we consider this epidemic spread model from a single source $I_0 = \{U\}$ on $\mathrm{NW}_n(\rho)$ with i.i.d. $\mathrm{Exp}(1)$ transmission times. We define

$$\mathrm{I}_n(t, U) := \frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{i \text{ is infected before or at time } t\} = \frac{1}{n} \#\{i : i \in [n], \Gamma_{U,j} \leq t\}, \quad (1.2)$$

the fraction of infected vertices at time t of the epidemic started from the vertex U.

Theorem 1.3 (Epidemic curve). *Let U be a uniformly chosen vertex in $[n]$, and let us consider the epidemic spread with source U and i.i.d. $\mathrm{Exp}(1)$ transmission times on $\mathrm{NW}_n(\rho)$. Then, the proportion of infected individuals satisfies w.h.p.*

$$\mathrm{I}_n(t + \tfrac{1}{\lambda} \log n, U) \xrightarrow{d} f(t + \tfrac{1}{\lambda} \log W_U),$$

where $f(t) = 1 - \mathrm{M}_{W_V}(x(t))$, where $M_{W_V}(\cdot)$ is the moment generating function of W_V, and $x(t) = \left(1 - \frac{1}{2}\pi_R^2\right) \mathrm{e}^{\lambda t}/(\lambda(\lambda + 1))$, with $\pi_R = 2/(\lambda + 2)$; and where W_U, W_V, are the same random variables as in Theorem 1.1.

Remark 1.4. The intuitive message of Theorem 1.3 is that a linear proportion of infected vertices can be observed after a time that is proportional to the logarithm of the size of the population. This time has a random shift given by $\frac{1}{\lambda} \log W_U$. Besides this random shift, the fraction of infected individuals follows a deterministic curve $f(\cdot)$: only the 'position of the curve' on the time-axes is random. A bigger value of W_U means that the local neighborhood of U is "dense", and hence the spread is quick in the initial stages: indeed, a bigger value of W_U shifts the function $f(t + (\log W_U)/\lambda)$ more to the left on the time axes. This phenomenon has been observed in real-life epidemics, see e.g. [2, 33] for a characterisation of typical epidemic curve shapes. For individual epidemic curves, browse e.g. [17].

The next proposition characterises the function $M_{W_V}(t)$ in the definition of the epidemic curve function $f(t)$ in Theorem 1.3.

Proposition 1.5 (Functional equation system for the moment generating function). *The moment generating function $M_{W_V}(\vartheta), \vartheta \in \mathbb{R}^+$ of the random variable W_V satisfies the following functional equation system, with $M_{W_V}(\vartheta) := M_{W^{(B)}}(\vartheta)$:*

$$M_{W^{(B)}}(\vartheta) = \left(\int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) \mathrm{e}^{-x} \mathrm{d}x \right)^2 \cdot \exp\left\{ \rho \cdot \int_0^\infty \left(M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) - 1 \right) \mathrm{e}^{-x} \mathrm{d}x \right\},$$

$$M_{W^{(R)}}(\vartheta) = \int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) \mathrm{e}^{-x} \mathrm{d}x \cdot \exp\left\{ \rho \cdot \int_0^\infty \left(M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) - 1 \right) \mathrm{e}^{-x} \mathrm{d}x \right\}.$$

$$(1.3)$$

Remark 1.6. These functional equations and the fact that there exists a solution for all $\vartheta \in \mathbb{R}^+$ follow from the usual branching recursion of multi-type branching processes, that can be found e.g. in [5].

1.3. **Related literature, comparison and context.** First passage percolation (FPP) was first introduced by Hammersley and Welsh [21] to study spreading dynamics on lattices, in particular on $\mathbb{Z}^d, d \geq 2$. The intuitive idea behind the method is that one imagines water flowing at a constant rate through the (random) medium, the waterfront representing the spread. The model turned out to be able to capture the core idea of several other processes, such as weighted graph distances and epidemic spreads.

Janson [24] studied typical distances and the corresponding hopcount, flooding times as well as diameter of FPP on the complete graph. He showed that typical distances, the flooding time and diameter converge to 1, 2, and 3 times $\log n/n$, respectively, while the hopcount is of order $\log n$.

Universality class. In a sequence of papers (e.g. [11, 12, 10, 23]) van der Hofstad *et al.* investigated FPP on random graphs. Their aim was to determine *universality classes* for the shortest path metric for weighted random graphs without 'extrinsic' geometry (e.g. the supercritical Erdős-Rényi random graph, the configuration model, or rank-1 inhomogeneous random graphs). They showed that typical distances and the hopcount scale as $\log n$, as long as the degree distribution has finite asymptotic variance and the edge weights are continuous on $[0, \infty)$. On the other hand, power-law degrees with infinite asymptotic variance drastically change the metric and there are several universality classes, compare [23] with [10]. In this respect, Theorems 1.1 and 1.2 show that the presence of the circle does not modify the universality class of the model.

Comparison to the Erdős-Rényi graph. Notice that the subgraph formed by shortcut edges is approximately an Erdős-Rényi graph, with the difference that the presence of the cycle always makes $\mathrm{NW}_n(\rho)$ connected and hence there is no subcritical or critical regime in $\mathrm{NW}_n(\rho)$. Typical distances on the Erdős-Rényi graph with parameter ρ/n and $\mathrm{Exp}(1)$ edge weights scale as $\log n/(\rho - 1)$ [11], while for $\mathrm{NW}_n(\rho)$ they scale as $(\log n)/\lambda$, with $\lambda = (\rho - 1 + \sqrt{\rho^2 + 6\rho + 1})/2 > \rho - 1$ for all $\rho \in \mathbb{R}$. This means that when $\rho > 1$, the presence of the cycle makes typical distances shorter, and this appears already in the constant scaling factor of $\log n$. However, $\lambda(\rho)/\rho \to 1$ as $\rho \to \infty$ meaning that the effect of the cycle becomes more and more negligible as the number of shortcut edges grow.

Comparison to inhomogeneous random graphs. Kolossváry *et al.* [27] studied FPP on the inhomogeneous random graph model (IHRG), defined in [14]. In this model, vertices have types from a type space S, and conditioned on the types of the vertices, edges are present independently with probabilities that depend on the types. One can fine-tune the parameters of this model so that any finite neighborhood of a vertex in the $\mathrm{NW}_n(\rho)$ model is similar to that of in the IHRG, that is, both of them can be modelled using *the same* continuous time multi-type branching process. It would be natural to conjecture that typical distances are then the same in these two models. It turns out that this is almost but not entirely the case: the first order term $\lambda^{-1} \log n$, and the random variables W_U, W_V are the same, but the additive constant c in Theorem 1.1 is not: the geometry of the Newman-Watts model modifies how the two branching processes can connect to each other, which modifies the constant. Writing the main result in [27] in the same form as the one in Theorem 1.1, we obtain $c_{\mathrm{IHRG}} = \log \left((\rho + 2)(2\rho + \lambda^2)/(\rho(\lambda + 2)^2 \lambda(\lambda + 1))\right)$.

The epidemic curve. In [13] Bhamidi *et al.* pointed out the connection between FPP, typical distances, and the epidemic curve by studying the epidemic spread on the configuration model with arbitrary continuous edge-weight distribution. Earlier, [8] Barbour and Reinert investigated the epidemic curve on the Erdős-Rényi random graph and on the configuration model with bounded degrees, where also possible other aspects such as contagious period of vertices or dependence of the transmission time distribution on the degrees might be present.

Possible future directions. In [3, 9, 18] the competition of two spreading processes running on the same graph is investigated. This can be considered a competition between two epidemics, as well as the word-of-mouth marketing of two similar products. The results suggest that the outcome depends on the universality class of the model: in ultra-small worlds, one competitor only gets a negligible part of the vertices, while on regular graphs coexistence might be possible, i.e., both colors can paint a linear fraction of vertices. Studying competition on $\mathrm{NW}_n(\rho)$ is an interesting and challenging future project.

1.4. **Structure of the paper.** In what follows, we prove Theorems 1.1, 1.2 and 1.3. The brief idea of the proof is the following: we choose two vertices uniformly at random, then we start to explore the neighbourhoods of these vertices in the graph in terms of the distance from these vertices (Section 2). We show that this procedure w.h.p. results in 'shortest weight trees' (SWT's) that can be coupled to two independent copies of a continuous time multi-type branching process (CMBP). We then handle how these two shortest weight trees connect in the graph in Section 3 with the help of a Poisson approximation. We provide the proof of Theorem 1.3 about the epidemic curve in Section 4 based on our result on distances. Finally we prove the Central Limit Theorem for the hopcount in Section 5, based on an indicator representation of the 'generation of vertices' in the branching processes.

2. Exploration process

To explore the neighborhood of a vertex, we use a modification of Dijkstra's algorithm.

Introduce the following notations: $\mathcal{N}(t), \mathcal{A}(t), \mathcal{U}(t)$ denote the set of explored (dead), active (alive) and unexplored vertices at time t, respectively, and $\mathrm{N}(t), \mathrm{A}(t), \mathrm{U}(t)$ for the sizes of these sets. The remaining lifetime of some vertex $w \in \mathcal{A}(t)$ at time t is denoted by $\mathcal{R}_w(t)$, and means that w will become explored exactly at time $t + R_w(t)$. The set of remaining lifetimes is $\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$. As before, $\mathrm{H}(v)$ denotes the neighbors of a vertex v.

2.1. **The exploration process on an arbitrary weighted graph.** Let $i = 1$. The vertex from which we start the exploration process is denoted by v_1. We color v_1 blue and set the time as $t = T_1 = 0$. Evidently, we take

$$\mathcal{N}(0) = \{v_1\}, \quad \mathcal{A}(0) = \mathrm{H}(v_1), \quad \mathcal{U}(0) = [n] \setminus (\{v_1\} \cup \mathrm{H}(v_1)).$$

The remaining lifetimes are determined by the edge weights, i.e.

$$\mathcal{R}_{\{\mathcal{A}(0)\}}(0) = \{\mathcal{R}_w(0) = X_{(v_1,w)} \text{ for all } w \in \mathrm{H}(v_1)\}.$$

We color the active vertices $w \in \mathrm{H}(v_1)$ to have the same color as the edge (v_1, w).

We work with induction from now on. In each step, we increase i by 1. We can construct the continuous time process in steps, namely, at the random times when we explore a new vertex.

Let $\tau_i = \min\left(\mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1})\right)$, the minimum of remaining lifetimes. Then define $T_i := T_{i-1} + \tau_i$, the time when we explore the next vertex. Nothing changes in the time interval $[T_{i-1}, T_i)$, hence for any $t \in [T_{i-1}, T_i)$,

$$\mathcal{N}(t) := \mathcal{N}(T_{i-1}), \quad \mathcal{A}(t) := \mathcal{A}(T_{i-1}), \quad \mathcal{U}(t) := \mathcal{U}(T_{i-1}).$$

From all the remaining lifetimes, we subtract the time passed: for some $0 \le s \le \tau_i$,

$$\mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1} + s) := \mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_{i-1}) - s,$$

subtracted element-wise. At time T_i, the vertex (or all the vertices, if there is more than one such vertices) v_i of which the remaining lifetime equals 0, becomes explored and its

neighbors become active. We shall refer to v_i as the ith explored vertex. We set

$$\mathcal{N}(T_i) := \mathcal{N}(T_{i-1}) \cup \{v_i\}, \quad \mathcal{A}(T_i) := (\mathcal{A}(T_{i-1}) \setminus \{v_i\}) \cup \mathrm{H}(v_i), \quad \mathcal{U}(T_i) := \mathcal{U}(T_{i-1}) \setminus \mathrm{H}(v_i).$$

We refresh the set of remaining lifetimes:

$$\mathcal{R}_{\{\mathcal{A}(T_i)\}}(T_i) := \mathcal{R}_{\{\mathcal{A}(T_{i-1})\}}(T_i) \setminus \{\mathcal{R}_{v_i}(T_i)\} \cup \{\mathcal{R}_x(T_i) : x \in \mathrm{H}(v_i)\}$$

where $\mathcal{R}_x(T_i) = X_{(v_i,x)}$, the edge weight of (v_i, x), and x also gets the color of (v_i, x).
On an arbitrary connected weighted graph, the exploration process can be continued until
all vertices become explored. Note that this algorithm builds the shortest weight tree SWT
from the starting vertex. This tree will be modeled using the branching process.

Remark 2.1. The set of active vertices might contain several occurrences of a vertex, in
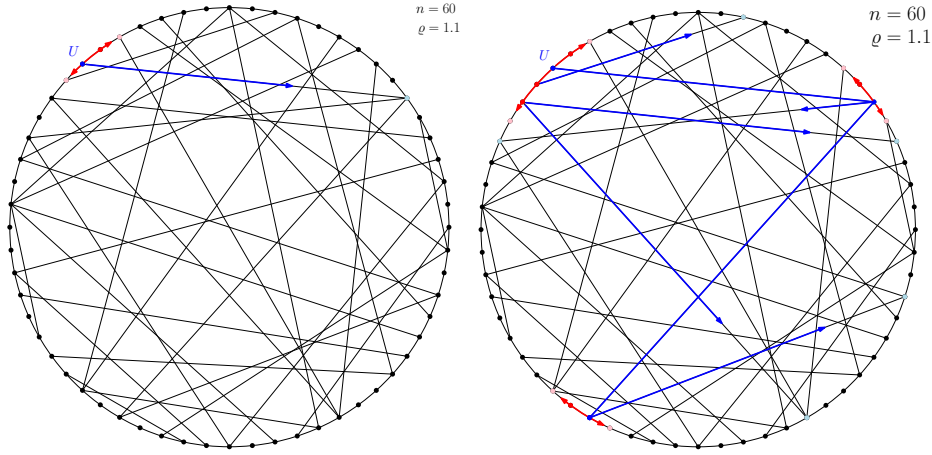case at least two neighbors of a vertex are explored already, see Figure 3.

FIGURE 1. A realisation of the Newman-Watts model for $k = 1$ and $\rho = 1.1$ with 60 vertices. On these two pictures, we illustrated the growing neighbourhood of a uniformly picked vertex. Circle edges are red and global edges are blue in the exploration. The edges that are partially red or blue are the ones that have an already explored vertex on one side while a not-yet explored (active) vertex on the other side.

2.2. **Exploration on the weighted Newman-Watts random graph.** Note that when
applying the exploration process above on a realization of $\mathrm{NW}_n(\rho)$, we can reveal the presence
of the edges and their weights in $\mathrm{NW}_n(\rho)$ along the exploration process. In this respect, all
the above quantities become random variables. Here we investigate the behaviour of this
random exploration process.

Let us color the cycle-edges red and the shortcut-edges blue, and let us say that a vertex is
red/blue if it is reached first via a red/blue edge during the exploration. We allow double (or
more) occurrences of the same vertex in $[n]$ among the active vertices (with two remaining
life-times along the two edges) in the exploration and also contradicting colorings. When such
a vertex gets explored for the first time, it gets the color that corresponds to the remaining
lifetime that became 0 (and forget about the other colors). Below, adding the subscript R

or B to any quantity corresponds to the same quantity restricted to only the red or blue vertices, respectively.

When running the exploration process, we build a weighted tree along the process containing the edges that are used to discover the new vertices in the exploration (this is a tree since we do not explore vertices twice). This tree has root v_1, grows in time, and at any time t it contains a vertex $v \in [n]$ precisely when $\mathcal{P}(v_1, v) < t$. Let us denote the tree up to time t by $\mathrm{SWT}^{v_1}(t)$.

Claim 2.2 (Children). *Suppose the vertex v is being explored for the first time (i.e., not "double-explored"). If v is red, one new red and* $\mathrm{Binomial}(n - 3, \frac{\rho}{n})$ *many new blue active vertices are born. If v is blue, two new red and* $\mathrm{Binomial}(n - 4, \frac{\rho}{n})$ *many new blue active vertices are born. The number of new blue active vertices is asymptotically* $\mathrm{Poi}(\rho)$ *in both cases. Further, at any time t, the elements of $\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$ are i.i.d.* $\mathrm{Exp}(1)$ *random variables, and the next explored vertex is chosen uniformly over the set of active vertices.*

Proof. On a cycle there are two vertices neighboring a vertex, hence, if v is red, then it has been reached from one of his neighbors. The other one is added to the new red active vertices. If v is blue then it has been reached via a shortcut edge and hence both of its neighbors on the cycle are added to the new red active vertices. Since there are $\mathrm{Bin}(n - 3, \rho/n)$ many shortcut edges from a vertex, this is also the distribution of new blue active vertices born when exploring a red vertex. For the exploration of a blue vertex, we reached this vertex via a blue edge, hence an additional $\mathrm{Bin}(n - 4, \rho/n)$ new active blue vertices. Clearly, by the convergence of binomial to Poisson distribution, each vertex has asymptotically $\mathrm{Poi}(\rho)$ many blue neighbours. The second statement follows from the fact that the edge weights are i.i.d. exponential random variables, which has the memoryless property. Finally, note that at any time, $\mathcal{R}_{\{\mathcal{A}(t)\}}(t)$ consists of i.i.d. exponential random variables, and the algorithm takes the minimum of these. Clearly, the minimum of finite many absolutely continuous random variables is unique almost surely, and uniform over the indices. $\square$

2.3. **Multi-type branching processes.** We define the following continuous time multi-type branching process (CMBP) that will correspond to the initial stages of $\mathrm{SWT}(t)$.

There are two particle types, red (R) and blue (B), and their lifetime is $\mathrm{Exp}(1)$, independent from everything else. Particles give birth upon their death. They leave behind offspring as in Claim 2.2: each particle has $\mathrm{Poi}(\rho)$ many blue offspring, red particles have one, while blue particles have two red children. Dead and alive particles will correspond to explored and active vertices, respectively. With this wording, for the number of alive and dead particles, we define

Definition 2.3. *We shall write* $\mathbf{A}(t) = (\mathrm{A}_R(t), \mathrm{A}_B(t))$ *for the number of alive particles of each type,* $\mathrm{A}(t)$ *standing for the total number of alive particles. Let* $\mathrm{N}(t) = \mathrm{N}_R(t) + \mathrm{N}_B(t)$, *where* $\mathrm{N}_q(t)$ *means the number of dead particles of type $q = R, B$. We assume the above quantities to be right-continuous. Superscripts $(R), (B)$ refer to the process started with a single particle of the given type.*

The exploration process corresponds to the process started with a single blue-type particle, which dies immediately.

2.3.1. *Literature on multi-type branching processes.* Here we restate the necessary theorems from [5] which we will use.

Definition 2.4 (Mean matrix). *Let* $\mathrm{M}(t) := \mathrm{M}_{r,q}(t) = \mathbb{E}[\mathrm{A}_q^{(r)}(t)]$, $(q, r = R, B)$ *the mean matrix, where* $\mathrm{A}_q^{(r)}(t)$ *is as defined above in Definition 2.3.*

It is not hard to see that $\mathrm{M}(t)$ satisfies the semigroup property $\mathrm{M}(t + s) = \mathrm{M}(t)\mathrm{M}(s)$ and the continuity condition $\lim_{t \to 0} \mathrm{M}(t) = \mathrm{I}$, where I denotes the identity matrix. As a result, we have:

Theorem 2.5 (Athreya-Ney)**.** *There exists an infinitesimal generator matrix* Q *such that* $\mathrm{M}(t) = \mathrm{e}^{\mathrm{Q}t}$, *where* $Q_{r,q} = a_r \mathbb{E}[D_q^{(r)}] - \delta_{r,q}$. *Here,* a_r *is the rate of dying for a particle of type* r, *(i.e., the parameter of its exponential lifetime),* D *is the number of offspring with the same sub-end superscript conventions as in Definition 2.3, and* $\delta_{r,q} = \mathbb{1}_{\{r=q\}}$ *(i.e.,* $\delta_{r,q} = 1$ *if and only if* $r = q$).

In our case,

$$\mathrm{Q} = \begin{pmatrix} 0 & \rho \\ 2 & \rho - 1 \end{pmatrix}$$

Eigenvalues and eigenvectors of the Q *matrix.* Using the characteristic polynomial, for $\rho \geq 0$, the maximal eigenvalue λ and the second eigenvalue λ_2 is given by

$$\lambda = \frac{\rho - 1 + \sqrt{\rho^2 + 6\rho + 1}}{2}, \quad \lambda_2 = \frac{\rho - 1 - \sqrt{\rho^2 + 6\rho + 1}}{2}. \tag{2.1}$$

The normalised left eigenvector π that satisfies $\pi \mathrm{Q} = \lambda \pi$ gives the stationary type-distribution:

$$\boldsymbol{\pi} = (\pi_R, \pi_B) = \left(\frac{2}{\lambda + 2}, \frac{\lambda}{\lambda + 2} \right). \tag{2.2}$$

We denote the right (column) eigenvector of Q by $\mathbf{u}$ and normalize it so that $\boldsymbol{\pi}\mathbf{u} = 1$. For later use, without computing, we denote by $\mathbf{v}_2$ and $\mathbf{u}_2$ the left (row) and right (column) eigenvector of Q belonging to the eigenvalue λ_2. The most important theorem for our purposes is that the CMBP grows exponentially with rate λ (the so-called Malthusian parameter), more precisely,

Theorem 2.6 ([5])**.** *With the notation as above, almost surely,*

$$\lim_{t \to \infty} \mathbf{A}(t)\mathrm{e}^{-\lambda t} = W\boldsymbol{\pi}$$

where W *is a nonnegative random variable, the almost sure martingale limit of* $W_t := \mathbf{A}(t)\mathbf{u}\,\mathrm{e}^{-\lambda t}$. *Further,* $W > 0$ *almost surely on the event of non-extinction.*

Theorem 2.7 ([5])**.** *Define* T_m, *the* m^{th} *split time, as the time of the* m^{th} *death in the branching process. (We assume* $T_1 = 0$ *for the death of the root.) On the event* $\{W > 0\}$,

(i) *For each* $q \in (R, B)$, $\lim_{m \to \infty} \mathrm{N}_q(T_m)/\mathrm{N}(T_m) = \lim_{m \to \infty} \mathrm{N}_q(T_m)/m \overset{a.s.}{=} \pi_q$

(ii) $\lim_{m \to \infty} m\mathrm{e}^{-\lambda T_m} \overset{a.s.}{=} \frac{1}{\lambda}W$

Corollary 2.8. *For the vector of dead particles* $\mathbf{N}(t) = (N_R(t), N_B(t))$,

$$\mathbf{N}(t)\mathrm{e}^{-\lambda t} \overset{a.s.}{\longrightarrow} \frac{1}{\lambda}W\boldsymbol{\pi}.$$

Proof of Theorems 2.5, 2.6, 2.7 and Corollary 2.8. The proofs can be found in [5, Chapter V.7]. □

Throughout the next sections, we develop error bounds on the coupling between the branching process and the exploration process on the graph. For convenience, we introduce

$$t_n := \frac{1}{2\lambda} \log n, \tag{2.3}$$

the times we will observe the branching and exploration processes at, as well as

$$W^{(n)} := e^{-\lambda t_n} \mathrm{A}(t_n), \quad \text{with} \quad W^{(n)} \overset{a.s.}{\longrightarrow} W, \tag{2.4}$$

the approximations of the martingale limit W at the times t_n. Note that in our case, extinction can never occur, hence almost surely $W > 0$.

2.4. **Labeling, coupling, error terms.** In this section we develop a coupling between the CMBP discussed in the previous section and $\mathrm{SWT}(t)$, the exploration process on $\mathrm{NW}_n(\rho)$.

Error bound on coupling the offspring. The CMBP is defined with $\mathrm{Poi}(\rho)$ blue offspring distribution, while in the exploration process a vertex has $\mathrm{Bin}(n-3, \rho/n)$ or $\mathrm{Bin}(n-4, \rho/n))$ many blue children. Let $Y \sim \mathrm{Poi}(\rho)$ and $X \sim \mathrm{Bin}(n, \rho/n)$. By the usual coupling of binomial and Poisson random variables, $\mathbb{P}(X \neq Y) \leq \frac{\rho^2}{n}$. Let $Z \sim \mathrm{Bin}(n-3, \rho/n)$, $V \sim \mathrm{Bin}(3, \rho/n)$ independent. Then $Z \stackrel{\mathrm{d}}{=} X - V$, and under the usual coupling

$$\mathbb{P}(Z \neq Y) \leq \mathbb{P}(X \neq Y) + \mathbb{P}(V \neq 0) = \frac{\rho^2}{n} + \frac{3\rho}{n} + o(1/n^2).$$

For the blue offspring of a blue vertex $\hat{Z} \sim \mathrm{Bin}(n-4, \rho/n)$, by similar arguments $\mathbb{P}(\hat{Z} \neq Y) \leq \frac{\rho^2}{n} + \frac{4\rho}{n} + o(1/n^2)$ holds. Taking maximum and using union bound, the probability that up to k steps, at least one particle has different number of blue offspring in the exploration process and the Poisson branching process, is at most $k(\rho^2 + 4\rho)/n$.

2.4.1. *Labeling and thinning.* We relate the CMBP to the exploration process on $\mathrm{NW}_n(\rho)$ through the labeling of the earlier. Below, everything must be interpreted modulo n.
  (i) The root is labeled u, the source of the exploration process. u can be U, a uniformly chosen vertex in $[n]$.
  (ii) Every other particle gets a label when it is born.
  (iii) We distinguish "left type" and "right type" red children. Left type red particles have a left type red child, right type red particles have a right type red child, blue particles have a red child of both types.
  (iv) A left type red child of v gets label $v - 1$, a right type red child of v is labeled $v + 1$.
  (v) The blue children of v get a set of labels uniformly chosen from $[n]$.

Lemma 2.9. *We say that the labeling fails if two explored vertices share the same label (this still allows for several occurrences of the same label in the active set). The probability that the labeling fails at the ith split is at most $2i/n$.*

Proof. The labeling fails at the ith split if the splitting particle has a label that is already taken by an explored vertex. We distinguish two cases.
When a blue particle splits. Since the label of a blue particle is chosen uniformly in $[n]$, and there are at most $i - 1$ dead labels already, the probability that we choose from this set is $(i - 1)/n$.
When a red particle splits. Note that the labeling procedure ensures that whenever a blue particle v is explored, it starts a growing (possibly asymmetric) red interval of red vertices around it. A red vertex, upon dying, extends this interval in one direction (if it is left type, then towards the left). Note that the original vertex v in this interval had a uniformly chosen label in $[n]$. Let us denote the position of the kth explored blue vertex by c_k, and write $l_k(T_i)$ and $r_k(T_i)$ for the number of explored red vertices to the left and to the right of c_k after the ith split, $i \geq k$. Finally, we denote the whole interval of explored vertices around c_k after the ith split by $I_k(T_i)$. Recall that the process is by definition right-continuous.

In this setting, the label of a red vertex that is just being explored can coincide with the label of an already explored red vertex if and only if two intervals 'grow into each other' at the ith split. Denote by I^* the interval that grows at the ith split, write $c^*, r^*(T_{i-1}), l^*(T_{i-1})$

for the location of its blue vertex, right and left length, respectively. Then, $I^\star$ grows into another interval I_k if and only if c_k, the location of the blue vertex in I_k, is at position $c^* - l^*(T_{i-1}) - r_k(T_{i-1}) - 1$ or is at position $c^* + r^*(T_{i-1}) + l_k(T_{i-1}) + 1$. (The first case means that the furthest explored red vertex on the right of I_k was a red active child of the furthest explored left vertex in I^*). Since the location of c_k is uniform in $[n]$,

$$\mathbb{P}(I^*(T_{i-1}) \cap I_k(T_{i-1}) = \varnothing, I^*(T_i) \cap I_k(T_i) \neq \varnothing) = \frac{2}{n}.$$

Note that there are exactly as many intervals as blue explored vertices (at either T_{i-1} or T_i, since the ith explored vertex v_i must be red). Let the event $E_i = \{v_i$ is red and its label is already used$\}$. Hence,

$$\mathbb{P}(E_i) \leq \sum_{k=1}^{N_B(T_{i-1})-1} \frac{2}{n} = \frac{2}{n}\left(N_B(T_{i-1}) - 1\right) \leq \frac{2i}{n},$$

since there are at most i blue explored vertices. Note that the proof also applies when the new red explored vertex coincides with a formerly explored blue one, in case $l_k(T_{i-1}) = 0$ or $r_k(T_{i-1}) = 0$. Hence, the statement of the lemma follows. $\qquad\square$

In $\text{NW}_n(\rho)$, the shortest path (u, v) through x necessarily uses the shortest path between (u, x). As a result, in the CMBP, we also do not need later occurrences of the label x. Hence, we mark the second (or any later) occurrence of a label *thinned*, and all its descendants *ghosts*. We move towards bounding the proportion of ghosts among active individuals to carry on with the CMBP approximation. To determine whether a vertex is a ghost, we need knowledge about its ancestors.

2.4.2. *Ancestral line.* We approach the problem of ghost actives with the help of the ancestral line. We define the ancestral line $\text{AL}(w)$ of a vertex w as the chain of particles leading to w from the root, including the root and w itself. Then an alive particle is a ghost if and only if at least one of its ancestors is thinned. The ancestral line was introduced by Bühler in [15, 16] with the following observation: for each time interval $[T_k, T_{k+1})$ we can allocate a unique particle on the ancestral line that was active in the interval $[T_k, T_{k+1})$. For the following observations, we condition on $\{D_i, i = 1, ..., k\}$, where D_i is the total number of offspring of the ith splitting particle. Denote by G_k the generation of a uniformly chosen alive (active) particle W after the kth split. Then $G_k = L_1 + L_2 + ... + L_k$, where the indicators L_i are conditionally independent and $L_i = 1$ if and only if the ancestor of W that was alive in the time interval $[T_i, T_{i+1})$ was newborn (born at T_i). (A rewording of the indicators L_i is as follows: $L_i = 1$ if and only if the ith splitting particle is in $\text{AL}(W)$.)

Since W is chosen uniformly, and at each split the individual to split is also chosen uniformly among the currently active individuals, each one of these active individuals is equally likely to be an ancestor of V. Further, in the interval $[T_i, T_{i+1})$, D_i many particles are newborn, and S_i many are alive, which yields the probability $\mathbb{P}(L_i = 1 | D_i, i = 1, ..., k) = D_i/S_i$, see the discussion at the beginning of [16, Section 2.A]. We arrive to the following corollary:

Corollary 2.10. *The probability of the ith dying particle being an ancestor of W, a uniformly chosen active vertex after the kth split:*

$$\mathbb{P}(v_i \in \text{AL}(W) | D_i, i = 1, ..., k) = \mathbb{P}(L_i = 1) = \frac{D_i}{S_i}.$$

Expected proportion of thinned actives. Let us combine Corollary 2.10 and Lemma 2.9. To be able to do so, we need the following lemma. We will provide its proof later on.

Lemma 2.11. *For every $\varepsilon > 0$, there exists a positive integer-valued random variable $K = K(\varepsilon)$ so that K is always finite and for every $i > K$, $S_i = i\lambda(1 + o(i^{-1/2+\varepsilon}))$ holds.*

Recall that $t_n = (\log n)/2\lambda$, and it was chosen such that the number of active vertices is of order $\sqrt{n}$, and that $\mathcal{A}(t), \mathrm{A}(t), \mathcal{N}(t), \mathrm{N}(t)$ denotes the set and number of active and dead individuals in the CMBP at time t, respectively.

Lemma 2.12. *Let $\mathcal{A}_G(t) = \{w \in \mathcal{A}(t) : w$ is a ghost$\}$ the set of ghost active vertices at time t and $\mathrm{A}_G(t)$ its size. For every fixed $s \in \mathbb{R}$, the proportion $\mathrm{A}_G(t_n + s)/\mathrm{A}(t_n + s)$ tends to 0 in probability as n tends to infinity.*

Proof of Lemma 2.12. The proportion $\mathrm{A}_G(t)/\mathrm{A}(t) = \mathbb{P}(W \in \mathcal{A}_G(t))$, where W is uniform over $\mathcal{A}(t)$, i.e., uniformly chosen active individual. Recall that v_i is the particle that dies at T_i. For an event E, let us write $\mathbb{P}_k(E) := \mathbb{P}(E|D_i, i = 1, \ldots, k)$. Using these notation and Corollary 2.10 for the representation of the ancestral line of $V \in \mathcal{A}(t)$, we can write

$$\mathbb{P}_k(W \in \mathcal{A}_G(t)) \leq \sum_{i=1}^{\mathrm{N}(t)} \mathbb{P}_k(v_i \in \mathrm{AL}(W) \text{ and } v_i \text{ is thinned}),$$

Since the labeling is independent of the family tree,

$$\mathbb{P}_k(W \in \mathcal{A}_G(t)) \leq \sum_{i=1}^{\mathrm{N}(t)} \mathbb{P}_k(v_i \in \mathrm{AL}(W)) \cdot \mathbb{P}_k(v_i \text{ is thinned}) \leq \sum_{i=1}^{\mathrm{N}(t)} \frac{D_i}{S_i} \frac{2i}{n}. \qquad (2.5)$$

We apply Lemma 2.11 by splitting the sum for parts up to K and above, use $D_i < S_i$ for $i \leq K$:

$$\mathbb{P}(W \in \mathcal{A}_G(t)) \leq \sum_{i=1}^{K} \frac{2i}{n} + \sum_{i=K+1}^{\mathrm{N}(t)} \frac{D_i}{\lambda(1 + o(i^{-1/2+\varepsilon}))n}$$

$$\leq \frac{K^2}{n} + 2\frac{\sum_{i=K+1}^{\mathrm{N}(t)} D_i}{\lambda n} < \frac{K^2}{n} + 2\frac{\mathrm{A}(t) + \mathrm{N}(t)}{\lambda n}$$

where we used that all particles are either active or dead in the process and with a possible modification of K, we can have $(1 + o(i^{-1/2+\varepsilon}) > 1/2$ for all $i > K$. Next, we can use Corollary 2.8 and Theorem 2.6, which gives that $\mathrm{N}(t) + \mathrm{A}(t) = e^{\lambda t}(\frac{1}{\lambda} + 1)W^{(n)}(1 + o(1))$. Hence

$$\mathrm{A}_G(t_n + s)/\mathrm{A}(t_n + s) = \mathbb{P}(W \in \mathcal{A}_G(t_n + s)) \leq K^2/n + 2e^{\lambda(t_n+s)}\frac{\lambda + 1}{\lambda^2 n}W^{(n)}(1 + o(1)).$$

Setting $t_n = \log n/(2\lambda)$, the right hand side tends to 0 as $n \to \infty$, since $W^{(n)} \to W$ and K is a tight random variable (does not depend on n). $\qquad \square$

Let us now return to the proof of Lemma 2.11. This lemma follows from [4, Theorem 1, Theorem 2]. Here, we restate [4, Theorem 1] using our notations and for a special case, where each eigenvalue has multiplicity 1. This is sufficient for our purposes and easier than the general case.

Theorem 2.13 (Asmussen, [4])**.** *Let $\mathbf{Z}_n$ be the number of individuals in the nth generation of a (discreet time) supercritical multi type Galton-Watson process, with dominant eigenvalue λ, the corresponding left and right eigenvector $\mathbf{v}$ and $\mathbf{u}$. For any other eigenvalue ν, $\mathbf{v}_\nu$ and*

$\mathbf{u}_\nu$ *denote the left and right eigenvectors belonging to* ν.
For an arbitrary vector $\mathbf{a} \in \mathbb{R}^p$ *with the property* $\mathbf{v} \cdot \mathbf{a} = 0$ *define*

$$\mu := \sup\{\nu : \mathbf{v}_\nu \mathbf{a} \neq 0\}, \quad \sigma^2 := \lim_{n\to\infty} \frac{|\mathbf{v}|\text{Var}(\mathbf{Z}_n\mathbf{a})}{\lambda^n} \tag{2.6}$$

If $\mu^2 < \lambda$, *then with* $C_n = (2\sigma^2\mathbf{Z}_n\mathbf{u}\log n)^{1/2}$

$$\liminf_{n\to\infty} \frac{\mathbf{Z}_n\mathbf{a}}{C_n} = -1 \quad and \quad \limsup_{n\to\infty} \frac{\mathbf{Z}_n\mathbf{a}}{C_n} = 1.$$

We also restate [4, Theorem 2] without change.

Theorem 2.14 (Asmussen 2., [4]). *Replacing* $\mathbf{Z}_n$ *with* $\mathbf{A}(t), t \in [0,\infty)$, *Theorem 2.13 remains valid for any supercritical irreducible multi-type Markov branching process.*

Proof of Lemma 2.11. We use the previous two theorems for the 2-type branching process defined in Section 2.3. Since π and $\mathbf{v}_2$ are linearly independent, for any $\mathbf{a} \neq (0,0)$ with $\pi\mathbf{a} = 0$, necessarily $\mathbf{v}_2\mathbf{a} \neq 0$, which implies $\mu = \lambda_2$ in (2.6). The eigenvalues of the mean matrix $\text{M}(t)$ are $e^{\lambda t}$ and $e^{\lambda_2 t}$. The condition $\mu^2 < \lambda$ in Theorem 2.13 is then equivalent to $2\lambda_2 < \lambda$ which follows from the nonnegativity of ρ, through simple algebraic computations, see (2.1). The asymptotic variance σ^2 and C_t in this case becomes:

$$\sigma^2 = \lim_{t\to\infty} \pi\text{Var}(\mathbf{A}(t)\mathbf{a})e^{-\lambda t}, \quad C_t = (2\sigma^2\mathbf{A}(t)\mathbf{u}\log t)^{1/2}$$

This implies that the theorem rewrites to

$$\limsup_{t\to\infty} \frac{\mathbf{A}(t)\mathbf{a}}{C_t} = 1 \text{ and } \liminf_{t\to\infty} \frac{\mathbf{A}(t)\mathbf{a}}{C_t} = -1.$$

Applying this for the split time T_i, we get that there is only a finite number of indexes i such that $|\mathbf{A}(T_i)\mathbf{a}/C_{T_i}| > 2$. Let the maximum of these indexes be K, a random variable. Since $T_i - \log i/\lambda$ has an almost sure limit by Theorem 2.7, T_i is of order $\log i$. This implies that C_{T_i} is of order $(i\log\log i)^{1/2}$, and by definition of the almost sure convergence, C_{T_i} exceeds $i^{1/2+\varepsilon}$ only finitely many times for every $\varepsilon > 0$.

Since $\mathbb{E}[\mathbf{A}(t)\mathbf{a}] = 0$ if and only if $\pi\mathbf{a} = 0$, we can apply the theorem for the centered version $S_i^c := S_i - \mathbb{E}S_i$. Then for $i > K$, $|S_i^c| \leq C_{T_i}$. The fluctuation is of smaller order then S_i; itself, which means we can indeed write $S_i = i\lambda(1 + o(i^{-1/2+\varepsilon}))$. For more detail on this, see the proof of [25, Corollary 3.16]. $\qquad\square$

2.4.3. *The number of multiple active and active-explored labels.* Recall that both in the exploration process as well as in the branching process there might be multiple occurrences of active vertices, see Remark 2.1, as thinning only prevents multiple explored labels. Later we want to use that the number of different active labels that are not ghossts at T_i is approximately the same as S_i, i.e., there are not many multiple occurrences. In Lemma 2.12 we have seen that the proportion of ghosts is negligible on the time scale t_n, but we still have to deal with labels that are multiply active, or are explored and active at the same time. We will discuss these issues in the following five cases:

1. A blue active vertex has been already explored.
2. A red active vertex has been already explored.
3. A blue active vertex is also red active.
4. A vertex is double red active.
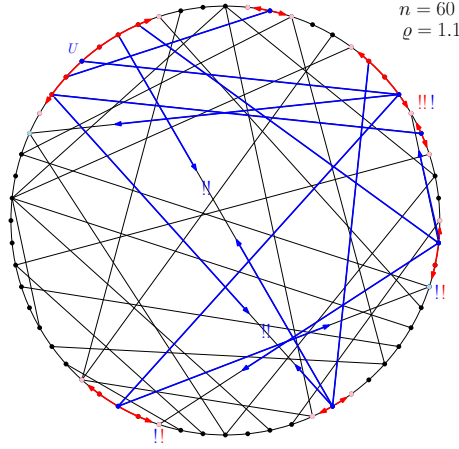5. A vertex is double blue active.

FIGURE 2. We indicated the growing neighbourhood of a uniformly picked vertex in the exploration process. Exclamation marks indicate 'bad events' for the coupling to a branching process: the vertices at the endpoint of edges (indicated along the edge) with two blue exclamation marks are vertices that are blue active and have been already explored as well. The vertex with two red and one blue exclamation mark is twice red active and once blue active at the same time.

We will denote by $p_\alpha(t)$ the probability that a uniformly chosen active vertex falls under case $\alpha = 1, \ldots, 5$ at time t, which is the same as the proportion of vertices falling under case α among all active vertices.

Case 1. *Blue active being already explored.* At time t, there are at most $N(t)$ explored labels that are not thinned. Under the condition that the active vertex is blue, its label is chosen uniformly over $[n]$, so the probability that this label has been already explored is at most $N(t)/n$. Substitute $N(t_n + s)$ from Corollary 2.8, then for $t_n + s = \frac{1}{2\lambda} \log n + s$,

$$p_1(t_n + s) \le \mathbb{P}(v \text{ is already explored} \mid v \text{ is blue}) = N(t_n + s)/n. \tag{2.7}$$

Case 2. *Red active being already explored.* This case can be treated similarly as the thinning of red vertices, so we also use the notation there. A label of the red active vertex is explored if and only if two intervals are about to grow into each other: the furthest explored red vertices in both intervals are neighbors. We call these intervals neighbors. Then, for two neighboring intervals, the active vertices at the end of each interval are explored in the other interval. Let I_k and $I_j, 1 \le k < j \le N_B(t)$ intervals with blue particles with label c_k and c_j respectively. Conditioned on $c_k, l_k(t), r_j(t)$, there are two possibilities so that I_k and I_j are neighbors: $c_j = c_i + r_i + l_j + 1$ or $c_j = c_i - l_i - r_j - 1$. Thus for each pair of indices the probability of the intervals being neighbors is $2/n$ (these are not independent, but expectation is linear). Summing up for all pair of indexes and dividing by the number of all red actives gives the proportion of *case 2* red actives among all red actives.

$$p_2(t_n + s) \le \frac{1}{A_B(t_n + s)} \sum_{1 \le i < j \le N_B(t_n+s)} \frac{2}{n} = \frac{\binom{N_B(t_n+s)}{2} \cdot 2/n}{2N_B(t_n + s)} \le \frac{N_B(t_n + s)}{2n}. \tag{2.8}$$

Case 3. *Blue active being red active.* Using that the labels of blue vertices are chosen uniformly,

$$\mathbb{P}(v \text{ is red and blue active}) = A_R(t_n + s)/n. \tag{2.9}$$

Case 4. *Multiple red active vertices.* This case is similar to Case 2. A vertex v can be red active twice if the two intervals that it belongs to are "almost neighbors", that is, both have v as an active vertex on one of their ends. (v is the only vertex separating them.) Conditioning on the location of one of the intervals, the blue vertex in the other interval can be at 2 different locations, hence

$$p_4(t_n + s) = p_2(t_n + s). \tag{2.10}$$

Case 5. *Multiple blue active vertices.* Again, the label of a blue vertex is chosen uniformly at random, hence the probability that the label of an active blue vertex v coincides with another active blue label is at most $A_B(t)/n$. Hence

$$p_5(t_n + s) \le \frac{1}{A_B(t_n + s)} \sum_{v \in \mathcal{A}_B(t_n+s)} A_B(t_n + s)/n = A_B(t_n + s)/n. \tag{2.11}$$

Corollary 2.15. *Define $A_e(t)$ the effective size of the active set as follows: we subtract from $A(t)$ the number of ghosts, already explored and multiple active labels, to get the number of different labels in $\mathcal{A}(t)$. Then*

$$A_e(t_n + s)/A(t_n + s) \xrightarrow{\mathbb{P}} 1.$$

Proof. By the previous arguments, a lower bound can be given if we subtract the individual probabilities for red and blue vertices to be deleted (note that this is a crude bound since we do not weight it with the proportion of red and blue active labels):

$$A_e(t_n + s)/A(t_n + s) \ge 1 - \sum_{i=1}^{5} p_i(t_n + s) = 1 - \frac{2N(t_n + s) + A(t_n + s)}{n},$$

where we summed up the rhs of (2.7), (2.8), (2.9), (2.10), (2.11) to obtain the rhs. Now we can use that $t_n + s = \log n/(2\lambda) + s$, use $N(t)$ from Corollary 2.8, and Theorem 2.6 to get

$$A_e(t_n + s)/A(t_n + s) \ge 1 - \frac{\lambda + 2}{\lambda} e^{\lambda s} W^{(n)}(1 + o(1))/\sqrt{n},$$

which tends to 1 since $W^{(n)} \to W$ a.s. by (2.4) and Theorem 2.6. $\square$

The conclusion of this section is summarized in the following corollary.

Corollary 2.16. *Fix $n \ge 1$ and $\rho > 0$. Consider the thinned CMBP with label u for the root. Then, there is a coupling of shortest weight tree $\mathrm{SWT}^u(t)$ in $\mathrm{NW}_n(\rho)$ to the evolution of the thinned CMBP as long as $t \le t_n + M$ for some arbitrary large $M \in \mathbb{R}$. Further, the set of active vertices in the thinned CMBP can be approximated by the set of labeled active vertices in the the original CMBP in the sense that the proportion of the different labels among the actives over the total active vertices tends to zero as $n \to \infty$, in the sense of Corollary 2.15.*

3. Connection process

Now that we have a good approximation of the shortest weight tree (SWT) started from a vertex, it provides us a method to observe the shortest weight path between two vertices. Let us give a raw sketch of this method before moving into the details. The previous section provides us with a coupling of a CMBP and the SWT as long as the total number of vertices is of order $\sqrt{n}$ in the SWT. To find the shortest weight path between vertices U and V, we grow the shortest weight trees from one of the vertices (SWT^U) until time t_n (the size is then of order $\sqrt{n}$). Then, conditioned on the presence of $\mathrm{SWT}^U(t_n)$, we grow SWT^V and see

when is the first time that these two trees intersect. The shortest weight path is determined by the first intersection of the *explored* set of vertices in the two processes. Note that a vertex w in the active set of vertices in $SWT^U(t_n)$ is at distance $t_n + R_w(t_n)$ from U. Since we have a good bound on the effective size of the set of active vertices, it turns out to be easier to look at the times when the first few active vertices in $\mathrm{SWT}^U(t_n)$ become explored in SWT^V, and then minimize over the total distance of the path formed by connecting U and V via this vertex. This is what we shall carry out now rigorously.

Definition 3.1 (Collision and connection)**.** *We grow* SWT^U*, the shortest weight tree of* U *until time* $t_n = \frac{1}{2\lambda} \log n$*, and then fix it. Then we grow* SWT^V *until time* $t_n + M$*, for some large* $M \in \mathbb{R}$*, conditioned on the presence of* $\mathrm{SWT}^U(t_n)$*. We say that a collision happens at time* $t_n + s$ *when an active vertex in* $\mathrm{SWT}^U(t_n)$ *becomes explored in* SWT^V *at time* $t_n + s$*. Denote the set of collision times by the point process* $(t_n + P_i)_{i \in \mathbb{N}}$*.*[1] *If a collision happens at vertex* x_i *at time* $t_n + P_i$*, this determines a path between* U *and* V *with length* $2t_n + P_i + \mathcal{R}^U_{x_i}(t_n)$*, where* $\mathcal{R}^U_{x_i}(t_n)$ *is the remaining lifetime of* x_i *in* $\mathrm{SWT}^U(t_n)$*. Then the length of the shortest weight path is given by*

$$\min_{i \in \mathbb{N}} \left(2t_n + P_i + \mathcal{R}^U_{x_i}(t_n + P_i) \right)$$

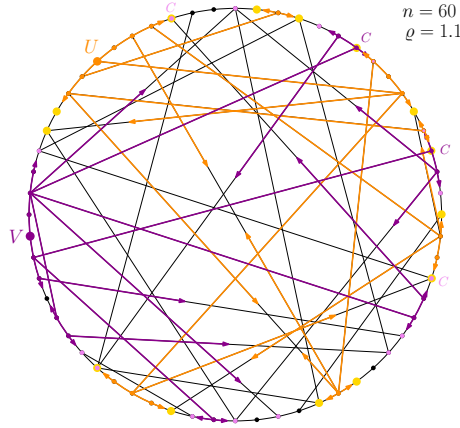*among all collision events.*

$n = 60$
$\varrho = 1.1$

FIGURE 3. We indicated the growing neighbourhood of two uniformly picked vertices in the exploration process, with purple and yellow colors, respectively. The letters 'C' on the picture show that a collision event happens at the given vertex. Notice that all these collision events have a remaining edge-length yet to be covered in the exploration process of U, i.e., the vertex is active in SWT^U and explored in SWT^V.

We can see that in the case of growing SWT^V after SWT^U, the labels belonging to explored vertices in SWT^U can not be used again, leading to some extra thinned vertices in SWT^V. We claim that the number of additional ghosts is not too big. (Since we would like to get a bound on the effective size of active vertices in $SWT^V(t_n + u)$, we must delete the descendants of vertices that formed earlier collision events.)

---

[1]We will see later that a.s. there is a first collision time. Hence, indexing by $i \in \mathbb{N}$ makes sense.

Claim 3.2. *Consider the case of growing* SWT^V *after* $\mathrm{SWT}^U(t_n)$ *on the same graph* $\mathrm{NW}_n(\rho)$. *Then the effective size of the active set in* SWT^V *for times* $t = t_n + s$ *is asymptotically the same as the size of the active set, that is, the statement of Corollary 2.15 remains valid for* SWT^V *as well.*

Note that it suffices to bound the proportion of ghosts, as the error terms caused by multiple active, or active-explored vertices are not increased by the presence of SWT^U.

Proof. We consider the computations in the proof of Lemma 2.12, using (2.5). Recall that the proportion of ghosts depends simultaneously on the thinning probability of the ith explored vertex as well as it being an ancestor of a uniform active vertex.

The arguments with the ancestral line (see 2.4.2) remain valid without any modification, we only have to examine the change in the thinning probability.

In case the ith explored vertex is blue, its label is chosen uniformly, thus the probability that this label coincides with a previously chosen label equals $\left(\mathrm{N}^U(t_n) + i - 1\right)/n$. In case the ith explored vertex in SWT^V is red, we can use the same idea as before: it has the label of an already explored vertex if and only if two intervals grow into each other with the ith step. We now consider the union of the intervals in SWT^U and SWT^V. Conditioned on the interval that grows, for any interval the probability that these two grow into each other is $2/n$. The number of intervals is at most the total number of blue explored vertices, $\mathrm{N}_B^U(t_n) + N_B^V(T_i)$. Hence the probability that the labeling fails at the ith step of SWT^V, if this is a red vertex is at most

$$\sum_{k=1}^{\mathrm{N}_B^U(t_n)+N_B^V(T_i)} \frac{2}{n} = \frac{2\left(\mathrm{N}_B^U(t_n) + N_B^V(T_i)\right)}{n} \leq \frac{2\left(\mathrm{N}^U(t_n) + i\right)}{n}.$$

Since the color of the i-th explored vertex is either blue or red, we get

$$\mathbb{P}(\text{ labeling fails in } \mathrm{SWT}^V \text{ at step i }) \leq 2\left(\mathrm{N}^U(t_n) + i\right)/n.$$

For the probability of a uniformly chosen active vertex in SWT^V being a ghost, similarly to (2.5), we have

$$\frac{\mathrm{A}_G^V(t_n + s)}{\mathrm{A}^V(t_n + s)} = \sum_{i=1}^{\mathrm{N}^V(t_n+s)} \frac{D_i}{S_i} \cdot \frac{2(\mathrm{N}^U(t_n) + i)}{n} = \sum_{i=1}^{\mathrm{N}^V(t_n+s)} \frac{D_i}{S_i} \cdot \frac{2i}{n} + \sum_{i=1}^{\mathrm{N}^V(t_n+s)} \frac{D_i}{S_i} \cdot \frac{2}{n} \mathrm{N}^U(t_n). \quad (3.1)$$

By Lemma 2.12, the first sum on the rhs tends to 0 as n tends to ∞. for the second sum, let us recall the a.s. finite K in Lemma 2.11, and we split the sum again. We use Corollary 2.8, 2.4 and $t_n = \log n/(2\lambda)$ to get

$$\frac{2}{n} \mathrm{N}^U(t_n) \sum_{i=1}^K \frac{D_i}{S_i} \leq \frac{2}{n} \mathrm{N}^U(t_n) \sum_{i=1}^K 1 = K \frac{2}{\lambda\sqrt{n}} W_U^{(n)} \overset{n\to\infty}{\longrightarrow} 0.$$

For the second part of the sum, by Lemma 2.11 again,

$$\Sigma_2 = \frac{2}{n} \mathrm{N}^U(t_n) \sum_{i=K+1}^{\mathrm{N}^V(t_n+s)} D_i/S_i = \frac{2}{n} \mathrm{N}^U(t_n) \sum_{i=K+1}^{\mathrm{N}^V(t_n+s)} \frac{D_i}{i\lambda(1 + o(i^{-1/2+\varepsilon}))}$$

Using $\mathbb{E}[D_i] \leq \lambda + 2$, we bound the expected value of the sum $\Sigma_3 := \sum_{i=K+1}^{\mathrm{N}^V(t_n+s)} \frac{D_i}{i\lambda(1+o(i^{-1/2+\varepsilon}))}$ with tower rule.

$$\mathbb{E}[\Sigma_3] \leq \mathbb{E}\left[(1+o(1)) \sum_{i=K+1}^{\mathrm{N}^V(t_n+s)} \frac{\lambda+2}{i\lambda}\right] \leq \frac{\lambda+2}{\lambda}\mathbb{E}\left[\log(\mathrm{N}^V(t_n+s))\right](1+o(1))$$

Since the logarithm is concave, we use Jensen's inequality:

$$\mathbb{E}[\Sigma_3] \leq \frac{\lambda+2}{\lambda} \log\mathbb{E}\left[\mathrm{N}^V(t_n+s)\right](1+o(1)) \tag{3.2}$$

From Theorem 2.5 it follows that $\mathbb{E}\left[\mathrm{N}^V(t_n+s)\right] = (0,1)\exp\{Q\cdot(t_n+s)\}\mathbf{1}$, where $\mathbf{1} = (1,1)^T$ is a column vector. Using the Jordan decomposition of the matrix Q and exponentiating, elementary matrix analysis yields that the leading order is determined by the main eigenvalue λ and hence $(1,0)\exp\{Q(t_n+s)\}\mathbf{1} \leq e^{\lambda(t_n+u)}C_1$ for some constant $C_1 \geq 0$. Let us then use this bound with $t_n + s = \log n/(2\lambda) + s$ to give an upper bound on the rhs of (3.2), and set $C_2 := 2C_1(\lambda+2)$. Then Markov's inequality yields

$$\mathbb{P}\left(\Sigma_3 \geq C_2(\log n)^2\right) \leq 1/\log n \stackrel{n\to\infty}{\longrightarrow} 0.$$

Then on the w.h.p. event $\{\Sigma_3 \leq C_2(\log n)^2\}$ for $\Sigma_2 = \frac{2}{n}\mathrm{N}^U(t_n)\Sigma_3$, by Corollary 2.8, (2.4) again,

$$\Sigma_2 \leq \frac{2}{n}\mathrm{N}^U(t_n)C_2(\log n)^2 \leq C_2(\log n)^2 e^{\lambda u}\frac{W_V^{(n)}(1+o(1))}{\lambda\sqrt{n}} \stackrel{n\to\infty}{\longrightarrow} 0.$$

$\square$

3.1. The Poisson point process of collisions.

Recall that we say that a collision event happens at time $t_n + s$ when an active vertex in $\mathrm{SWT}^U(t_n)$ becomes explored in SWT^V at time $t_n + s$. First we show that for each pair of colours, with respect to the parameter s in $t_n + s$, the set of points $(P_i)_{i\in\mathbb{N}}$ form a non-homogeneous Poisson point process (PPP) on $\mathbb{R}$, and that these PPP-s are asymptotically independent. We consider the intensity measure $\mu(\mathrm{d}t)$, $t \in \mathbb{R}$ as the derivative of the mean function $M(t)$ (expected value of points up till time t). To determine the intensity measure of the collision process, we will consider the four collision point processes for each possible pair of colours. None of the PPPs is empty: since the labels of blue vertices are chosen uniformly, they can meet any color, and considering the growing set of intervals, we see that red can meet red as well.

Let us introduce the notation for $q, r \in \{R, B\}$, $s \in \mathbb{R}$

$$\mathcal{C}_{q,r}(s) := \mathcal{N}_q^V(t_n+s) \cap \mathcal{A}_r^U(t_n), \quad \mathrm{C}_{q,r}(s) := |\mathcal{C}_{q,r}(s)|,$$
$$\mathcal{C}(s) := \bigcup_{q,r\in\{R,B\}} \mathcal{C}_{q,r}(s), \quad \mathrm{C}(s) := |\mathcal{C}(s)| \tag{3.3}$$

(Note that e.g. $\mathcal{C}_{R,B}(s)$ denotes the set of *red* explored labels in SWT^V that are *blue* active in SWT^U.) The corresponding intensity measures are denoted by $\mu_{q,r}(s)$ and total intensity measure by $\mu(s)$.

In Corollary 2.15 we showed that the effective size of the active set at times $t_n + s$ for $s \in \mathbb{R}$ is asymptotically the same as the number of active individuals in the CMBP, so we can use the asymptotics of $(A_R(t), A_B(t))$ from Theorem 2.6. We shall handle the cases when blue vertices are involved in a similar fashion. For the coming sections, for $s \in \mathbb{R}$ and and event E we define the notation $\mathbb{P}_s(E) := \mathbb{P}(E|\mathrm{A}_B^U(t_n), \mathrm{A}^U R(t_n), \mathrm{N}_B^V(t_n+s), \mathrm{N}_R^V(t_n+s))$.

Blue-blue collision. By the definition of the set $\mathcal{C}_{B,B}(s)$, we can use the following indicator representation:

$$\mathrm{C}_{B,B}(s) = \sum_{x \in \mathcal{N}_B^V(t_n+s)} \mathbb{1}\{x \in \mathcal{A}_B^U(t_n)\}. \tag{3.4}$$

Recall that the labels in $\mathcal{A}_B^U(t_n)$ are chosen uniformly in $[n]$, hence for any label $x \in [n]$,

$$\mathbb{P}_s(x \in \mathcal{A}_B^U(t_n)) = 1 - (1 - 1/n)^{\mathrm{A}_B^U(t_n)} = \mathrm{A}_B^U(t_n)/n(1 + o(1)). \tag{3.5}$$

Further, since $\mathrm{A}_B^U(t_n) = O(\sqrt{n})$, elementary calculation using inclusion-exclusion formula shows that the indicators in (3.4) are weakly dependent in the sense that $\mathbb{P}_s(y \in \mathcal{A}_B^U(t_n)|x \in \mathcal{A}_B^U(t_n)) = \mathrm{A}_B^U(t_n)/n(1 + o(1))$. Further, the events $\{x \in \mathcal{N}_B^V(t_n + s)\}, \{x \in \mathcal{A}_B^U(t_n)\}$ are independent, the usual Poisson approximation for weakly dependent random variables (e.g. via factorial moments, as in [22, Theorems 2.4, 2.5]) yields that $\mathrm{C}_{B,B}(s)$ converges to a Poisson variable for each $s \in \mathbb{R}$, and by Wald's equation,

$$\mathbb{E}_s[\mathrm{C}_{B,B}(s)] = \mathrm{N}_B^V(t_n + u) \cdot \mathrm{A}_B^U(t_n)/n(1 + o(1)).$$

By noting that $\mathbb{1}\{x \in \mathcal{N}_B^V(t_n + s_1)\} \leq \mathbb{1}\{x \in \mathcal{N}_B^V(t_n + s_2)\}$ when $s_1 \leq s_2$, we also get that $\mathrm{C}_{B,B}(s_1) \leq \mathrm{C}_{B,B}(s_2)$ and hence by standard methods one can show that the process $(\mathrm{C}_{B,B}(s))_{s \in \mathbb{R}}$ converges to a Poisson point process in $\mathbb{R}$.

Red-blue collision. Using the same argument as for the previous case,

$$\mathrm{C}_{R,B}(s) = \sum_{x \in \mathcal{N}_R^V(t_n+s)} \mathbb{1}\{x \in \mathcal{A}_B^U(t_n)\}, \quad \mathbb{E}_s[\mathrm{C}_{R,B}(s)] = \mathrm{N}_R^V(t_n + s) \cdot \mathrm{A}_B^U(t_n)(1 + o(1)).$$

Further, for $y \in \mathcal{N}_R^V(t_n+s)$ and $x \in \mathcal{N}_R^V(t_n+s)$, the indicators $\mathbb{1}\{z \in \mathcal{A}_B^U(t_n)\}$ for $z = x, y$ are only weakly dependent, we can also conclude that the process $(\mathrm{C}_{R,B}(s))_{s \in \mathbb{R}}$ is asymptotically independent of $(\mathrm{C}_{B,B}(s))_{s \in \mathbb{R}}$.

Blue-red collision. This time we approach the number of collision events through indicators of labels in the explored set, i.e.,

$$\mathcal{C}_{B,R}(s) = \mathcal{N}_B^V(t_n + s) \cap \mathcal{A}_R^U(t_n), \quad \mathrm{C}_{B,R}(u) = \sum_{x \in \mathcal{A}_R^U(t_n)} \mathbb{1}\{x \in \mathcal{N}_B^V(t_n + s)\}. \tag{3.6}$$

The blue labels in $\mathcal{N}_B^V(t_n + s)$ are chosen uniformly in $[n]$, hence, a similar calculation as in (3.5), and the independence of the labeling in $\mathcal{N}_B^V(t_n + s)$ and in $\mathcal{A}_R^U(t_n)$ yields again

$$\mathbb{E}_s[\mathrm{C}_{R,B}(s)] = \mathrm{A}_R^U(t_n) \cdot \mathrm{N}_B^V(t_n + s)/n(1 + o(1)).$$

The independence of the three processes $\left(\mathrm{C}_{R,B}(s), \mathrm{C}_{R,B}(s), (\mathrm{C}_{B,R}(s))\right)_{s \in \mathbb{R}}$ can be seen by 'reversing' the indicators in (3.4) to get a similar form as the one in (3.6), and noting that the variables included are again only weakly dependent.

Using the asymptotic results for $\mathbf{A}(t), \mathbf{N}(t)$ (from Theorem 2.6 and Corollary 2.8) and the definition of $W^{(n)}$ in (2.4), then differentiating with respect to s yields

$$\mu_{B,B}(\mathrm{d}s) = \pi_B^2 W_V^{(n)} W_U^{(n)} \mathrm{e}^{\lambda s} \mathrm{d}s(1 + o(1)),$$
$$\mu_{R,B}(\mathrm{d}s) = \pi_B \pi_R W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda s} \mathrm{d}s(1 + o(1)), \tag{3.7}$$
$$\mu_{B,R}(\mathrm{d}s) = \pi_B \pi_R W_V^{(n)} W_U^{(n)} \mathrm{e}^{\lambda s} \mathrm{d}s(1 + o(1)),$$

where the $1 + o(1)$ factor only depends on n and comes from the error of the approximation $N_q(t) = \mathrm{e}^{\lambda t} \lambda^{-1} W^{(n)} \pi_q (1 + o(1))$. It is not hard to show (e.g. using Borel-Cantelli lemma) that these PPP-s have only finitely many points on $(-\infty, 0)$, hence, indexing the points by $i \in \mathbb{N}$ is doable.

Red-red collision. The red-red collision events have to be treated slightly differently, since here the geometry of the cycle plays a more important role. Recall that we stopped the evolution of SWT^U at time t_n. Hence, we consider $\mathrm{SWT}^U(t_n)$ as a fixed set of intervals, $\{I_k, k = 1, ..., \mathrm{N}_B^U(t_n)\}$ (some of them might have already possibly merged by time t_n), while the set of intervals $\{J_i(s), i = 1, ..., \mathrm{N}_B^V(t_n + s)\}$ in $\mathrm{SWT}^V(t_n + s)$ is growing (and possibly merging) with s. A collision happens when one of the intervals $J_i(s)$ grows into one of the intervals I_k.

Note that here we face a new technical issue. When two intervals $J_i(s)$ and I_k collide at time s, in principle we should stop the evolution of $J_i(s)$, that is, for all $s' > s$ we should have $J_i(s') \equiv J_i(s)$. But, since this would cause computational difficulties later, (since then we have to condition on all the earlier collisions to be able to calculate the intensity of the next one). Hence, it is easier to do the following approximation on the number of red-red collisions: we let $J_i(s)$ grow further and it might collide with more vertices inside I_k. The error terms caused by such events is negligible, since such events have been already treated when we investigated the 'extra' thinning of SWT^V imposed by $\mathrm{SWT}^U(t_n)$, that had a negligible contribution in the sense of Claim 3.2.

We decompose the number of collisions as a sum of indicators. Recall the notation from paragraph 2.4.1: c_k, l_k, r_k stands for the location of the blue vertex, the number of red explored to the left and right from c_k in the interval I_k. We adapt the same notation for the intervals J_i by adding a superscript V to the above quantities. We claim that for any pair I_k and $J_i(s)$, the probability that they had already made a red-red collision by time $t_n + s$ is

$$\{J_i(s) \cap I_k \neq \varnothing \text{ via red-red collision}\} = \cup_{m=1}^{l_i(s)} \mathbb{1}\{c_i^V = c_k - l_k - 1 - m\}$$
$$\bigcup \cup_{m=1}^{r_i(s)} \mathbb{1}\{c_i^V = c_k + r_k + 1 + m\}. \tag{3.8}$$

To see this, we condition on c_k, r_k, l_k. Note that there is two active red vertices at the boundary of I_k: one on the left, at location $c_k - l_k - 1$ and one on the right, at location $c_k + r_k + 1$. When the one at location $c_k - l_k - 1$ coincides with any of the $r_i^V(s)$ explored right-type red vertices of $J_i(s)$, a (potentially past) collision has happened. Each of these events uniquely determines the position of c_i^V, the location of the blue vertex of $J_i(s)$, forming an interval of length $r_i^V(s)$ for potential locations. We remark that this left active vertex at location $c_k - l_k - 1$ can *not* coincide with a left-type red vertex in $J_i(s)$, since that would mean that either c_i^V also coincides with I_k – causing an event that we already counted in Claim 3.2, or, c_i^V is to the left of the whole interval I_k, and $l_i^V(s)$ is already so large that it swallowed the whole interval I_k, and this case also have been already treated in Claim 3.2. (Since in this case, the part of $J_i(s)$ that the overlaps with I_k has been already thinned.) Similarly, there are $l_i^V(s)$ many positions for c_i^V on the left of c_k that cause a valid collision event. Using that c_i^V is chosen uniformly over $[n]$ – since it is a blue vertex – yields the right hand side of (3.8). Note that these indicators are mutually exclusive, and hence

$$\mathbb{P}_s(J_i(s) \cap I_k \neq \varnothing \text{ via red-red collision}) = (l_i^V(s) + r_i^V(s))/n. \tag{3.9}$$

The number of red-red collisions can be obtained by summing up the right hand side of (3.8) for all intervals I_k and $J_i(s)$. This yields a collection of indicators that have a weak dependence structure: each indicators only depends on at most $\max_i |J_i(s)|$ many other indicators. Since $t = O(\log n)$, and the length of each interval is determined by consecutive exponential variables (i.e. the time to reach the k-th red vertex on the right of a blue vertex has a Gamma$(k, 1)$ distribution), it is elementary to show that $\max_i |J_i(s)| = o((\log n)^2)$ asymptotically almost surely. As a result, a Poisson approximation can be carried out by using e.g. the factorial moments of these indicators, in a similar fashion that the one in [22,

Proposition 7.12], using [22, Theorems 2.4, 2.5]. We leave the details to the reader. We get that $C_{R,R}(s)$ converges to a PPP with intensity measure given by

$$\mathbb{E}_s[C_{R,R}(s)] = \sum_{k=1}^{N_B^U(t_n)} \sum_{i=1}^{N_B^V(t_n+s)} (l_i^V(s) + r_i^V(s))/n.$$

Note that the inner sum is simply $N_R^V(t_n + s)$, while $N_B^U(t_n) = A_R^U(t_n)/2$, since there are two red actives in each interval I_k. Then,

$$\mathbb{E}_s[C_{R,R}(s)] = A_R^U(t_n) \cdot N_R^V(t_n + s)/(2n).$$

Further, similar arguments as before show that this PPP is asymptotically independent of $C_{B,B}(s), C_{B,R}(s), C_{R,B}(s)$, and we also obtain

$$\mu_{R,R}(\mathrm{d}s) = \frac{1}{2}\pi_R^2 W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda s}\mathrm{d}s(1 + o(1)). \tag{3.10}$$

We emphasize that to get (3.7) and (3.10) we assumed that the number of actual intervals in $\mathrm{SWT}^U(t_n)$ and $\mathrm{SWT}^V(t_n + s)$ is $N_B^U(t_n)$ and $N_B^V(t_n + s)$, respectively. This is not entirely true due to the fact that intervals within SWT^U or within SWT^V might have merged already earlier. However, in this case some of the included vertices are ghosts: Corollary 2.15 showes that the effective size of the active set at times $t_n + s$ for $s \in \mathbb{R}$ is asymptotically the same as the number of active individuals in the CMBP, which, by the fact that every interval has precisely two active red vertices, implies that also the number of *disjoint* active intervals is asymptotically the same as $N_B^U(t_n)$ and $N_B^V(t_n + s)$, respectively in the two processes. We have arrived to the following theorem:

Theorem 3.3 (Total intensity measure of the collision PPP). *The total intensity measure of the collision Poisson point process is*

$$\mu(\mathrm{d}s) = W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda u}\big(\pi_B^2 + 2\pi_B\pi_R + \pi_R^2/2\big) = W_U^{(n)} W_V^{(n)} \mathrm{e}^{\lambda s}\big(1 - \pi_R^2/2\big)\mathrm{d}s(1+o(1)), \tag{3.11}$$

where the factor $(1 + o(1))$ only depends on n.

Proof. We have calculated the individual intensity measures of the four Poisson point processes in (3.7) and (3.10). These processes are asymptotically independent, since the total collection of *all* the indicator variables included in their construction are only weakly dependent. One way then is to show that the joint factorial moments converge, in a similar manner than the one in [22, Proposition 7.12]. Or, one might also consider the point process of the total intensities immediately together to get (3.11). It is important to observe that the depletion of the number of active individuals due to earlier collisions does not influence the asymptotic intensities, since the first is of constant order on the time scale $t = t_n + s, s \in \mathbb{R}$, while the latter is of order $\sqrt{n}$, the size of the active/explored labels in each of the colors. ∎

3.2. **Proof of Theorem 1.1.** It is well known [32] that if $(E_i)_{i\in N}$ is a collection of i.i.d. random variables with distribution function $F_E(y)$, and the points $(P_i)_{i\in\mathbb{N}}$ form a one-dimensional Poisson point process with intensity measure $\mu(\mathrm{d}s)$ on $\mathbb{R}$, then the points $(P_i, E_i)_{i\in\mathbb{N}}$ form a two-dimensional non-homogeneous Poisson point process on $\mathbb{R} \times \mathbb{R}$, with intensity measure $\mu(\mathrm{d}s) \times F_E(\mathrm{d}y)$.

In our case, to get the shortest path between U and V, recall from Definition 3.1 that we have to minimize the sum of time of the collision and the remaining lifetimes over the collision events. Mathematically, we want to minimize the quantity $P_i + E_i$ over all points
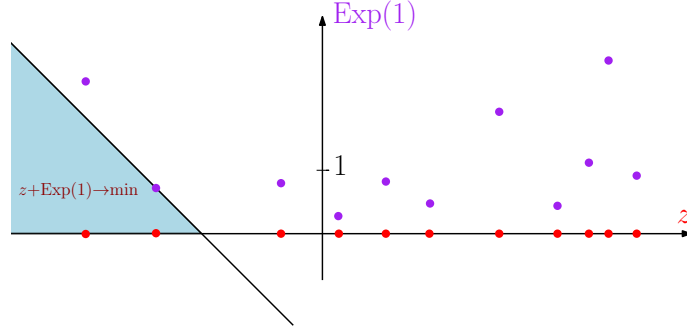
FIGURE 4. An illustration of the two dimensional PPP with points $(P_i, E_i)_{i \in \mathbb{N}}$, with the point minimizing the sum $P_i + E_i$ indicated.

(P_i, E_i) of the two dimensional PPP with intensity measure $\nu(\mathrm{d}s \times \mathrm{d}y) := \mu(\mathrm{d}s) \times \mathrm{e}^y \mathrm{d}y$, since the remaining lifetimes are i.i.d. exponential random variables.

Note that event $\{\min_i P_i + E_i \geq z\}$ is equivalent to the event that there is no point in the infinite triangle $\Delta(z) = \{(x, y) : y > 0, x + y < z\}$ in this two-dimensional PPP. We calculate

$$\nu(\Delta(z)) = \int_{-\infty}^{z} \int_{0}^{z-s} \mu(\mathrm{d}s) \cdot \mathrm{e}^{-y} \mathrm{d}y = W_U^{(n)} W_V^{(n)} \left(1 - \frac{1}{2} \pi_R^2 \right) \frac{1}{\lambda(\lambda + 1)} \mathrm{e}^{\lambda z} (1 + o(1)).$$

For short, we introduce

$$\mathcal{W}^{(n)} := W_U^{(n)} W_V^{(n)} \left(1 - \frac{1}{2} \pi_R^2 \right) \frac{1}{\lambda(\lambda + 1)} (1 + o(1)), \tag{3.12}$$

Then we can reformulate

$$\nu(\Delta(z)) = \mathcal{W}^{(n)} \mathrm{e}^{\lambda z} \tag{3.13}$$

Let us turn our attention back to $\mathcal{P}_n(U, V)$, the shortest weight path between U and V. By the previous argument, we conclude

$$\mathbb{P}(\mathcal{P}_n(U, V) \geq z + 2t_n | \mathcal{W}^{(n)}) = \mathbb{P}(\mathrm{Poi}(\nu(\Delta(z))) = 0 | \mathcal{W}^{(n)}) = \exp\{-\nu(\Delta(z))\} \tag{3.14}$$

Rearranging the left hand side and substituting the computed value of $\nu(\Delta(z))$, we get

$$\mathbb{P}(2t_n - \mathcal{P}_n(U, V) \leq -z | \mathcal{W}^{(n)}) = \exp\{-\mathcal{W}^{(n)} \mathrm{e}^{\lambda z}\}$$

We substitute $t_n = \log n/(2\lambda)$, and set $z := -x/\lambda$ to get

$$\mathbb{P}\left(\log n - \lambda \mathcal{P}_n(U, V) < x | \mathcal{W}^{(n)}\right) = \exp\{-\exp\{-(x - \log \mathcal{W}^{(n)})\}\}$$

We recognize on the right hand side the cumulative distribution function of a shifted Gumbel random variable, which implies

$$(\log n - \lambda \mathcal{P}_N(U, V)) | \mathcal{W}^{(n)} \stackrel{d}{=} \Lambda + \log \mathcal{W}^{(n)},$$

Rearranging and substituting $\mathcal{W}^{(n)}$ from (3.12), and using that the martingales $(W_U^{(n)}, W_V^{(n)}) \stackrel{a.s.}{\longrightarrow} (W_U, W_V)$,

$$\mathcal{P}_n(U, V) - \frac{1}{\lambda} \log n \stackrel{d}{\longrightarrow} = -\frac{1}{\lambda} \Lambda - \frac{1}{\lambda} \log(W_U W_V) - \frac{1}{\lambda} \log \left(1 - \frac{1}{2} \pi_R^2 \right) + \frac{1}{\lambda} \log (\lambda(\lambda + 1)).$$

This finishes the proof of Theorem 1.1.

4. Epidemic curve

Recall the definition of the epidemic curve function from Section 1.2. The discussion of the epidemic curve consists of three parts: first, we find the correct function f by computing the first moment of $\mathrm{I}_n(t, U)$. Then we prove the convergence in probability by bounding the second moment. Finally, we give a characterization of M_{W_V}, the moment generating function of the random variable W_V, that determines the epidemic curve function f.

4.1. **First moment.** First, we condition on the value $W_U^{(n)}$ from the martingale approximation of the branching process of the uniformly chosen vertex U. Then we can express the fraction of infected individuals as a sum of indicators, and calculate its conditional expectation:

$$\mathbb{E}\left[\mathrm{I}_n(t, U)\,\big|\,W_U^{(n)}\right] = \frac{1}{n}\sum_{w\in[n]} \mathbb{P}\left(w \text{ is infected by time t}\,\big|\,W_U^{(n)}\right).$$

Note that the rhs equals the probability of a *uniformly chosen* vertex, which we shall denote by V, being infected. Also note that a vertex is infected if and only if its distance from U is shorter than the time passed, hence

$$\mathbb{E}\left[\mathrm{I}_n(t, U)\,\big|\,W_U^{(n)}\right] = \mathbb{P}\left(\mathcal{P}_n(U, V) \leq t\,\big|\,W_U^{(n)}\right). \tag{4.1}$$

Now we can further condition on $W_V^{(n)}$ and use the distribution of $\mathcal{P}_n(U, V)$ conditioned on $W_U^{(n)}, W_V^{(n)}$. Recall from Subsection 3.2, that

$$\mathbb{P}\left(\mathcal{P}_n(U, V) \geq z + 2t_n\,\big|\,W_U^{(n)}, W_V^{(n)}\right) = \exp\left\{-W_U^{(n)} W_V^{(n)} \left(1 - \tfrac{1}{2}\pi_R^2\right) \tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda z}\right\}. \tag{4.2}$$

Let us set here $z = t - \frac{1}{\lambda}\log W_U^{(n)}$ and rearrange, yielding

$$\mathbb{P}\left(\mathcal{P}_n(U, V) \leq t - \tfrac{1}{\lambda}\log W_U^{(n)} + \tfrac{1}{\lambda}\log n\,\big|\,W_U^{(n)}, W_V^{(n)}\right)$$
$$= 1 - \exp\left\{-W_V^{(n)} \left(1 - \tfrac{1}{2}\pi_R^2\right) \tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\right\}$$

Then, from (4.1) and (4.3) we get

$$\mathbb{E}\left[\mathrm{I}_n(t - \tfrac{1}{\lambda}\log W_U^{(n)} + \tfrac{1}{\lambda}\log n, U)|W_U^{(n)}\right] = 1 - \mathbb{E}\left[\exp\left\{-W_V^{(n)}\left(1 - \tfrac{1}{2}\pi_R^2\right)\tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}\right\}\right]. \tag{4.3}$$

We recognize that the second term on the right hand side is the moment generating function of $W_V^{(n)}$, $\mathrm{M}_{W_V^{(n)}}(x)$, at $x(t) = -\left(1 - \tfrac{1}{2}\pi_R^2\right)\tfrac{1}{\lambda(\lambda+1)}\mathrm{e}^{\lambda t}$.

Changing variables yields

$$\mathbb{E}[\mathrm{I}_n(t + \tfrac{1}{\lambda}\log n, U)|W_U^{(n)}] = 1 - M_{W_V^{(n)}}\left(x(t + \tfrac{1}{\lambda}\log W_U^{(n)})\right).$$

Note that $W_V^{(n)}$ converges to W_V almost surely, which implies their moment generating functions converge in probability. This function is exactly the one given in Theorem 1.3.

4.2. **Second moment.** The first moment above showed that the expected value of $\mathrm{I}_n(t, U)$ indeed converges in probability to the defined f function at the given point. We prove Theorem 1.3 by showing that the variance of $\mathrm{I}_n(t, U)$ converges to 0, then Chebyshev's inequality yields that $\mathrm{I}_n(t, U)$ converges to its expectation in probability.

Denote by $\mathbb{1}_i = \mathbb{1}\{i \text{ is infected by time } t\}$. Let us calculate

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_i|W_U^{(n)}\right) = \frac{1}{n^2}\sum_{i\in[n]}\mathrm{Var}(\mathbb{1}_i|W_U^{(n)}) + \frac{2}{n^2}\sum_{i<j\in[n]}\mathrm{Cov}[\mathbb{1}_i, \mathbb{1}_j|W_U^{(n)}]$$

Since $\mathbb{1}_i$ is an indicator, $\mathrm{Var}(\mathbb{1}_i | W_U^{(n)}) \leq 1$, hence the first term on the rhs is at most $\frac{1}{n}$. As for the second term,

$$\mathrm{Cov}[\mathbb{1}_i, \mathbb{1}_j | W_U^{(n)}] = \mathbb{E}[\mathbb{1}_i \mathbb{1}_j | W_U^{(n)}] - \mathbb{E}[\mathbb{1}_i | W_U^{(n)}] \mathbb{E}[\mathbb{1}_j | W_U^{(n)}]$$

$$= \mathbb{P}(i \text{ and } j \text{ are both are infected} | W_U^{(n)}) - \mathbb{P}(i \text{ is infected} | W_U^{(n)}) \mathbb{P}(j \text{ is infected} | W_U^{(n)})$$

Imagine now three exploration processes on NW_n, one from U, one from i and one from j. It is not hard to see that the three exploration processes from these three vertices can be approximated by three *independent* branching processes. This implies that the covariance can be bounded by the error of coupling between the graph and the branching processes, as well as the thinning inside one tree and between the trees: these all have error terms of order at most $1/\log n$. It is not hard to see that the coupling can be extended to three SWT's (instead of two, as before), and the error terms increase only by constant multiples. The connection processes between SWT^U and the other two are related only through the intersection of SWT^i and SWT^j, which is again at most of order $1/\log n$. As a result, then $\mathbb{P}(i \text{ and } j \text{ are both are infected} | W_U^{(n)}) - \mathbb{P}(i \text{ is infected} | W_U^{(n)}) \mathbb{P}(j \text{ is infected} | W_U^{(n)}) = O(1/\log n)$.

This coupling works if i and j are fairly apart, say $(i-j) \bmod n > (\log n)^{1+\varepsilon}$ for any $\varepsilon > 0$ (this is w.h.p. longer than the length of the longest red interval). The number of "bad pairs", which are closer than this is $n(\log n)^{1+\varepsilon}/2$, compared to the number of all pairs $\binom{n}{2}$, the fraction goes to 0. Even for these, the covariance is bounded by 1. Then the sum divided by n^2 goes to 0.

With that, we have bounded the variance by a term that goes to 0, which finishes the proof.

4.3. **Characterization of the epidemic curve function.** In this section, we prove Proposition 1.5. Recall that adding a superscript (B) or (R) indicates a branching process described in Section 2.3 that is started from a blue or red type vertex, respectively. We start with the recursive formula for the martingale limit random variables from [5]:

$$W^{(B)} \overset{\mathrm{d}}{=} \sum_{i=1}^{D_R^{(B)}} \mathrm{e}^{-\lambda X_i} W_i^{(R)} + \sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda X_j} W_j^{(B)},$$

where $W_i^{(R)}$ are independent copies of $W^{(R)} = \lim_{t\to\infty} \mathrm{e}^{-\lambda t} Z^{(R)}(t)$, and $W_j^{(B)}$ are independent copies of $W^{(B)} \overset{\mathrm{d}}{=} W_V$, and X_i, X_j are i.i.d. $\mathrm{Exp}(1)$. Denote the moment generating functions of $W^{(B)}$ and $W^{(R)}$ by $M_{W^{(B)}}, M_{W^{(R)}}$ respectively. Recall that a blue individual has two red and $\mathrm{Poi}(\rho)$ many blue children. Hence

$$\mathbb{E}\left[\mathrm{e}^{\vartheta W^{(B)}}\right] = \left(\mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda X_i} W^{(R)}\}\right]\right)^2 \cdot \mathbb{E}\left[\exp\left\{\vartheta \sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda X_j} W_j^{(B)}\right\}\right] \qquad (4.4)$$

We use law of total expectation with respect to X_i to compute

$$J^{(R)} := \mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda X_i} W^{(R)}\}\right] = \int_0^\infty \mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda x} W^{(R)}\}\right] \mathrm{e}^{-x} \mathrm{d}x$$

$$= \int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) \mathrm{e}^{-x} \mathrm{d}x$$

Let $J^{(B)}$ defined similarly, with $M_{W^{(R)}}$ replaced by $M_{W^{(B)}}$. Then, the second factor in (4.4) can be treated by conditioning on $D_B^{(B)}$ and using independence:

$$\prod_{j=1}^{D_B^{(B)}} \mathbb{E}\left[\exp\{\vartheta \mathrm{e}^{-\lambda X_j} W_j^{(B)}\}\right] := \prod_{j=1}^{D_B^{(B)}} J^{(B)} = \left(J^{(B)}\right)^{D_B^{(B)}}.$$

Taking expectation w.r.t. $D_B^{(B)} \overset{\mathrm{d}}{=} \mathrm{Poi}(\rho)$ yields that

$$\mathbb{E}\left[\exp\left\{\vartheta \sum_{j=1}^{D_B^{(B)}} \mathrm{e}^{-\lambda X_j} W_j^{(B)}\right\}\right] = \exp\left\{\rho\big(J^{(B)} - 1\big)\right\}$$

We can rewrite the factor in exponent as

$$J^{(B)} - 1 = \int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x - 1 = \int_0^\infty \left(M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) - 1\right)\mathrm{e}^{-x}\mathrm{d}x,$$

then the moment generating function in (4.4) becomes

$$M_{W^{(B)}}(\vartheta) = \left(\int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x\right)^2 \cdot \exp\left\{\rho \cdot \int_0^\infty \left(M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) - 1\right)\mathrm{e}^{-x}\mathrm{d}x\right\}.$$

Similarly for $M_W^{(R)}$, using that $D_R^{(R)} = 1$ and $D_B^{(R)} \overset{\mathrm{d}}{=} \mathrm{Poi}(\rho)$,

$$M_{W^{(R)}}(\vartheta) = \int_0^\infty M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x})\mathrm{e}^{-x}\mathrm{d}x \cdot \exp\left\{\rho \cdot \int_0^\infty \left(M_{W^{(R)}}(\vartheta \mathrm{e}^{-\lambda x}) - 1\right)\mathrm{e}^{-x}\mathrm{d}x\right\}$$

We have just showed that the moment generating functions satisfy the system of equations given in Proposition 1.5, and by [5], there exist proper moment generating functions satisfying these functional equations.

5. Central limit theorem for the hopcount

In this section we prove Theorem 1.2 that states that the hopcount $\mathrm{H}_n(U, V)$, the number of edges along the shortest weight path between two vertices U and V chosen uniformly at random, satisfies a central limit theorem with mean and variance both $\frac{\lambda+1}{\lambda} \log n$.

For this, we consider the shortest weight path between U and V in two parts: the path from U within $\mathrm{SWT}^U(t_n)$ and from V within SWT^V, to the vertex where the connection happens. We denote the vertex where the connection happens by Y. These paths are disjoint with the exception of Y, hence if suffices to determine their lengths, i.e. the graph distance of Y from U and Y from V. Denote by $G^{(U)}(Y)$ the generation of Y in SWT^U, similarly for V. Then the required steps from the root U to Y is exactly $G^{(U)}(Y)$.

Claim 5.1. *The choice of Y is asymptotically independent in the two* SWT*'s.*

Proof. Conditioned on Y being the connecting vertex, it is uniformly chosen over the active set of SWT^V. That determines its label, and it determines which particle is chosen in SWT^U *through the label*. Since the labeling is independent of the structure of the family tree, aside from the thinning, the choice of Y in SWT^U is independent from its choice in SWT^V. We have already bounded the fraction of ghost particles (those who have one of their ancestors thinned) by a term that goes to 0 in Lemma 2.12, hence asymptotic independence holds. $\square$

With these notations, $\mathrm{H}_n(U, V) = G^{(U)}(Y) + G^{(V)}(Y)$, and the two terms are independent. We reformulate the theorem using these terms:

$$\frac{\mathrm{H}_n(U, V) - \frac{\lambda+1}{\lambda} \log n}{\sqrt{\frac{\lambda+1}{\lambda} \log n}} = \frac{G^{(U)}(Y) - \frac{\lambda+1}{2\lambda} \log n}{\sqrt{\frac{\lambda+1}{\lambda} \log n}} + \frac{G^{(V)}(Y) - \frac{\lambda+1}{2\lambda} \log n}{\sqrt{\frac{\lambda+1}{\lambda} \log n}} \qquad (5.1)$$

Considering that the terms are independent, it suffices to show that both terms on the right hand side have normal distribution with mean 0 and variance $\frac{1}{2}$. We show that both terms on the rhs of (5.1), multiplied by $\sqrt{2}$, have standard normal distribution. Due to the method we established the connection between SWT^U and SWT^V, the two terms need to be treated somewhat differently.

5.1. **Generation of the connecting vertex in** SWT^V. Recall that we established the connection between SWT^U and SWT^V in the following way: we grew SWT^U until time t_n, then we freeze its evolution. Then, we grow SWT^V, and every time a label is assigned to a splitting particle, we check if this label belongs to the active set of SWT^U. As a result, the connecting vertex Y is a particle at some splitting time T_k, and hence chosen uniformly over the active vertices. This implies that we can use for its generation the indicator decomposition of the ancestral line described in Section 2.4.2, for Y's generation as $G_k = \sum_{i=1}^{k} \mathbb{1}_i$, where conditioned on the offspring variables D_i, the indicators are independent and have success probability $\mathbb{P}(\mathbb{1}_i = 1) = \frac{D_i}{S_i}$.

In our case the number of splits is a random variable. Recall from Section 3.2 that the connection time minus t_n forms a tight random variable, (see e.g. (3.14)), hence till the connection there are $\mathrm{N}(t_n + Z)$ many explored vertices for some random $Z \in \mathbb{R}$. By Corollary 2.8, $\mathrm{N}(t_n + u) = C\sqrt{n}$ for some bounded random variable C (that might depend on n, but is tight). Denote by

$$B_1 = \frac{\sum_{i=1}^{C\sqrt{n}} \mathbb{1}_i - \sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}}{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}}, \quad B_2 = \frac{\sqrt{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i}\left(1 - \frac{D_i}{S_i}\right)}}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}},$$

$$B_3 = \frac{\sum_{i=1}^{C\sqrt{n}} \frac{D_i}{S_i} - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}}$$

$$(5.2)$$

Then

$$\frac{G^{(V)}(Y) - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = \frac{\sum_{i=1}^{C\sqrt{n}} \mathbb{1}_i - \frac{\lambda+1}{2\lambda}\log n}{\sqrt{\frac{\lambda+1}{2\lambda}\log n}} = B_1 \cdot B_2 + B_3 \tag{5.3}$$

Our aim is to show that Lindeberg's CLT is applicable for B_1, B_2 converges to 1, and B_3 converges to 0.

5.1.1. *Term B_1*. For this sum of (conditionally independent) indicators, Lindeberg's condition is trivially satisfied if the total variance tends to infinity. To give a lower bound,

$$\sum_{i=1}^{C\sqrt{n}} D_i/S_i \left(1 - D_i/S_i\right) = \sum_{i=1}^{C\sqrt{n}} D_i/S_i - \sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2. \tag{5.4}$$

Recall Lemma 2.11, and split the sum according to the random variable K. Each vertex has at least one red child, hence $D_i \geq 1$. Then

$$\sum_{i=1}^{C\sqrt{n}} D_i/S_i \geq \sum_{i=K+1}^{C\sqrt{n}} \frac{1}{i\lambda(1 + o(i^{-1/2+\varepsilon}))}.$$

where K is a.s. finite. The ith term on the rhs is at least $1/(2i)$, and thus the rhs tends to infinity, and is at least $\log n/(2\lambda)$. For the second term in (5.4), we can use that the second moment of $D_i \overset{d}{=} \mathrm{Poi}(\rho) + 1 + \mathbb{1}\{i^{\mathrm{th}} \text{ explored is blue}\}$ can be bounded by some constant M_2 independent of i. Hence, again cutting the sum at K, the sum of the first K terms is a.s. finite. For the rest, we can use Lemma 2.11 again, and then Markov's inequality yields:

$$\mathbb{P}\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i^2}{(i^2\lambda^2(1 + o(i^{-1/2+\varepsilon})^2)} \geq \sum_{i=K+1}^{C\sqrt{n}} \frac{M_2}{(i^2 2\lambda^2)} \cdot \log\log n\right) \leq \frac{1}{\log\log n}. \tag{5.5}$$

Note that $\sum_{i=1}^{\infty} M^2/(i^2 2\lambda^2) \leq \frac{M_2 \pi^2}{12}$. Combining the two estimates for the two terms in (5.4), we see that the variance tends to infinity w.h.p. As a result, the term B_1 in (5.2) satisfies a CLT.

5.1.2. *Term* B_2. Similarly as for the term B_1, we cut the sum at K given by Lemma 2.11 and write

$$\frac{\sum_{i=1}^{C\sqrt{n}} D_i/S_i(1-D_i/S_i)}{\frac{\lambda+1}{2\lambda}\log n} = \frac{\sum_{i=1}^{K} D_i/S_i(1-D_i/S_i)}{\frac{\lambda+1}{2\lambda}\log n} + \frac{\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i}{\frac{\lambda+1}{2\lambda}\log n} - \frac{\sum_{i=1}^{C\sqrt{n}} D_i^2/S_i^2}{\frac{\lambda+1}{2\lambda}\log n}.$$

The first fraction tends to 0, as the numerator is a.s. finite. For the numerator in the third term, we can use (5.5) again, which shows that the third term tends to zero w.h.p. We have yet to show that the second term tends to 1. Let $\mathcal{F}_n = \sigma(D_1, ..., D_n)$ be the filtration generated by the random variables D_i. Then

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} D_i/S_i}{\frac{\lambda+1}{2\lambda}\log n} = \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i-\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}}{\frac{\lambda+1}{2\lambda}\log n} + \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}}{\frac{\lambda+1}{2\lambda}\log n}. \tag{5.6}$$

For the first term of the rhs of (5.6), we will use Chebyshev's inequality. For this, elementary calculation using tower rule yields that

$$\mathrm{Var}\left[\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}\right] \leq \sum_{i,K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i^2|\mathcal{F}_{i-1}]}{\lambda^2 i^2(1+o(i^{-1/2+\varepsilon})^2}$$

Since $\mathbb{E}[D_i^2|\mathcal{F}_{i-1}] \leq M_2$ as in Section 5.1.1, we get that the rhs is at most $M_2 \pi^2/(3\lambda^2)$. Then Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}\right| \geq \log\log n \cdot \frac{\pi^2}{3}\frac{M_2}{\lambda^2}\right) \leq \frac{1}{(\log\log n)^2}. \tag{5.7}$$

This implies that the first term in (5.6) tends to 0 w.h.p.

Now to show that the second term in (5.6) tends to 1, we use a corollary of Theorem 2.6 (see [5]), stating that the vector $\left(\frac{S_{i-1}^R}{S_{i-1}}, \frac{S_{i-1}^B}{S_{i-1}}\right) \to (\pi_R, \pi_B)$ a.s. Further analysis (in particular, the central limit theorem about (S_i^R, S_i^B) in [25]) yields that the error term is at most of order $i^{-1/2+\varepsilon}$. Hence, using that $D_i \overset{d}{=} \mathrm{Poi}(\rho)+1+\mathbb{1}\{i^{\mathrm{th}}$ explored is blue$\}$ and the definition of λ it is elementary to show that

$$\mathbb{E}[D_i|\mathcal{F}_{i-1}] = (\lambda+1)(1+o(i^{-1/2+\varepsilon}))$$

Substituting this into the sum, we have

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}}{\frac{\lambda+1}{2\lambda}\log n} = \frac{\sum_{i=K+1}^{C\sqrt{n}} 1/i}{\log n/2} + \frac{\sum_{i=K+1}^{C\sqrt{n}} o(i^{-1/2+\varepsilon})/i}{\log n/2} \tag{5.8}$$

The first term on the rhs, introducing a constant error term δ from the integral approximation, equals

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} 1/i}{\log n/2} = \frac{\log(C\sqrt{n}) - \log(K+1) + \delta}{\log n/2} \overset{n\to\infty}{\longrightarrow} 1, \tag{5.9}$$

since C is a tight random variable. The second term in (5.8) is at most $\sum_{i=0}^{\infty} O(i^{-3/2+\varepsilon})$ is summable and finite, hence divided by $\log n$ it tends to 0. Combining everything, we get that B_2 in (5.2) tends to 1 w.h.p.

5.1.3. *Term B_3.* As before, we cut this sum at K given by Lemma 2.11, and the sum of the first K terms divided by $\log n$ tends to 0 since $D_i/S_i < 1$. When we consider the rest of the sum, we use the approximation of S_i (given by Lemma 2.11) and add and subtract $\mathbb{E}[D_i|\mathcal{F}_{i-1}]$ again:

$$\frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i}{S_i} - \frac{\lambda+1}{2\lambda} \log n}{\sqrt{\frac{\lambda+1}{2\lambda} \log n}} = \frac{\sum_{i=K+1}^{C\sqrt{n}} \frac{D_i - \mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}}{\sqrt{\frac{\lambda+1}{2\lambda} \log n}} + \frac{\left(\sum_{i=K+1}^{C\sqrt{n}} \frac{\mathbb{E}[D_i|\mathcal{F}_{i-1}]}{i\lambda(1+o(i^{-1/2+\varepsilon}))}\right) - \frac{\lambda+1}{2\lambda} \log n}{\sqrt{\frac{\lambda+1}{2\lambda} \log n}}$$

(5.10)

The numerator of the first term on the rhs has been treated in (5.7) and is w.h.p of order at most $\log \log n$, hence the first term on the rhs tends to 0 w.h.p. For the second term on the rhs of (5.10) we can use (5.8) and (5.9), and then it is at most

$$\frac{\lambda+1}{\lambda} \cdot \frac{\log C - \log(K+1) + \delta + \sum_{i=K+1}^{C\sqrt{n}} o(i^{-1/2+\varepsilon})/i}{\sqrt{\frac{\lambda+1}{2\lambda} \log n}} \to 0$$

almost surely, since C is a tight random variable. This shows that the term B_3 in (5.2) tends to 0 w.h.p, and finishes the proof of the CLT for the generation of the connecting vertex in SWT^V, see (5.3).

5.2. **Generation of the connecting vertex in SWT^U.** For the generation of Y in SWT^U, we have to use a different approach. This is because the label of the connecting vertex is chosen uniformly among the active vertex of SWT^V but is not necessarily uniform over the active vertices in SWT^U. Indeed, it is a longish but elementary calculation to show that conditioned on the event that a connection happens, any active red label in SWT^U is chosen with asymptotic probability $(\mathrm{A}^{(U)}(t_n))^{-1}(1 - \frac{\pi_R}{2})/(1 - \frac{\pi_R^2}{2})$, while any active blue label is chosen with asymptotical probability $(\mathrm{A}^{(U)}(t_n))^{-1}1/(1 - \frac{\pi_R^2}{2})$, where $\mathrm{A}^{(U)}(t_n)$ is the total number of active vertices in SWT^U. However, the following claim is still valid and will be enough to show the needed CLT:

Claim 5.2. *Conditioned on the connecting vertex having a label of a certain color in SWT^U, with high probability, it is chosen uniformly at random among the active labels of that color in SWT^U.*

Proof. We show the statement first for color blue. Recall that a blue label was chosen uniformly in $[n]$. Since the restriction of a uniform distribution to any set is again uniform, the probability of connection is the same for any particular blue active label among the *different* labels in the active blue labels in SWT^U. Recall that the number of different labels (that are neither thinned nor ghost) is called the effective size and it is treated in Corollary 2.15.

The problem here is though, that some labels in the branching process approximation are multiply active, and these are neither thinned nor called ghosts, and if chosen, they modify the uniform probability for the connection.[2] However, Corollary 2.15 implies that the fraction of multiple actives tends to 0 at time t_n. Hence if we pick an active label in $\mathrm{A}_B^U(t_n)$ it has multiplicity 1 w.h.p., including red and blue instances. (This also implies

---

[2]Consider a label V_i in the BP of SWT^U that is active m_i times, i.e., there are m_i individuals in the BP having label V_i. The label V_i in SWT^V is still chosen only with the same probability, (that is, $1/n$). Since which one of the m_i individuals has the minimal remaining lifetime is uniformly distributed, every individual with label V_i has probability $(m_i \mathrm{A}_{B,e}^{(U)}(t_n))^{-1}$ to be the connecting vertex, conditioned on connection at a blue label (with $\mathrm{A}_{B,e}^{(U)}$ being the effective size of blue labels).

that asymptotically, the label has a well defined color.) Hence with high probability, at the connecting vertex, we have a uniform distribution over all possible blue active labels.

An analogous argument can be carried through for red active labels as well, using the fact that the centre of the interval where they belong to is chosen uniformly, and the fact that multiple red labels have proportion tending to 0 at time t_n. $\qquad\square$

To finish the central limit theorem of $G_U(Y)$, we use a general result of Kharlamov [26] about the generation of a uniformly chosen active individual in a given type-set in a multi-type branching process. For this, consider a type set $\mathcal{S}$ of a multi-type branching process, and let $\mathcal{A}_\mathcal{S} = \cup_{q \in \mathcal{S}} \mathcal{A}_q$ the set of active individuals with any type from the type-set S. Then, [26, Theorem 2] states that the generation of a uniformly chosen individual in $\mathcal{A}_\mathcal{S}$ satisfies a central limit theorem with asymptotic mean and variance that is *independent* of the choice of $\mathcal{S}$.[3]

To apply this result, first pick $\mathcal{S} := \{R, B\}$ in our case. Then, the statement simply turns into a CLT of the generation of a uniformly picked active individual. We have seen when treating $G^{(V)}(Y)$ that the asymptotic mean and variance are both $\frac{\lambda+1}{2\lambda} \log n$ in this case.

Now apply the result again for $\mathcal{S} := \{R\}$ and $\mathcal{S} := \{B\}$, separately. Combined with the previous observation, we get that an individual chosen uniformly at random with color blue/red, respectively, also satisfies a CLT with the same asymptotic mean and variance. This, combined with Claim 5.2, implies that whether Y is red or blue in SWT^U, its generation $G^{(U)}(Y)$ admits a central limit theorem with mean and variance $\frac{\lambda+1}{2\lambda} \log n$. This completes the proof of Theorem 1.2.

References

[1] L. Addario-Berry and T. Lei. The mixing time of the Newman-Watts small world. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 1661–1668. SIAM, 2012.

[2] AFMC Primer on Population Health. Patterns of disease development in a population: the epidemic curve. http://phprimer.afmc.ca/Part2-MethodsStudyingHealth/Chapter7ApplicationsOfResearchMethodsInSurveillanceAndProgrammeEvaluation/Patternsofdiseasedevelopmentinapopulationtheepidemiccurve. Accessed 26 Feb 2015.

[3] T. Antunović, Y. Dekel, E. Mossel, and Y. Peres. Competing first passage percolation on random regular graphs. *ArXiv e-prints*, Sept. 2011.

[4] S. Asmussen. Almost sure behavior of linear functionals of supercritical branching processes. *Transactions of the American Mathematical Society*, 231(1):233–248, 1977.

[5] K. Athreya and P. Ney. *Branching Processes*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer Berlin Heidelberg, 1972.

[6] A. D. Barbour and G. Reinert. Small worlds. *Random Structures & Algorithms*, 19(1):54–74, 2001.

[7] A. D. Barbour and G. Reinert. Discrete small world networks. *Electron. J. Probab.*, 11:no. 47, 1234–1283, 2006.

[8] A. D. Barbour and G. Reinert. Approximating the epidemic curve. *Electron. J. Probab*, 18(54):1–30, 2013.

[9] E. Baroni, R. v. d. Hofstad, and J. Komjáthy. Fixed speed competition on the configuration model with infinite variance degrees: unequal speeds. arXiv:1408.0475 math.PR, 2014.

[10] S. Bhamidi, R. v. d. Hofstad, and G. Hooghiemstra. First passage percolation on random graphs with finite mean degrees. *The Annals of Applied Probability*, 20(5):1907–1965, 2010.

[11] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on the ErdősRényi random graph. *Combinatorics, Probability and Computing*, 20:683–707, 9 2011.

[12] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. Universality for first passage percolation on sparse random graphs. *ArXiv e-prints*, Oct. 2012.

---

[3]In fact, Kharlamov also specifies the asymptotic mean and variance, but that is for us given by another method, so for us this weak implication of [26, Theorem 2] is sufficient.

[13] S. Bhamidi, R. van der Hofstad, and J. Komjáthy. The front of the epidemic spread and first passage percolation. *J. Appl. Probab.*, 51A:101–121, 12 2014.

[14] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[15] W. Bühler. Generations and degree of relationship in supercritical markov branching processes. *Probability Theory and Related Fields*, 18(2):141–152, 1971.

[16] W. J. Bühler et al. The distribution of generations and other aspects of the family structure of branching processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*. The Regents of the University of California, 1972.

[17] Centers for Disease Control and Prevention. Reports of salmonella outbreak investigations from 2013. http://www.cdc.gov/salmonella/outbreaks-2013.html. Accessed 26 Feb 2015.

[18] M. Deijfen and R. van der Hofstad. The winner takes it all. *ArXiv e-prints*, June 2013.

[19] R. Durrett. *Random graph dynamics*, volume 20. Cambridge university press, 2007.

[20] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[21] J. M. Hammersley and D. J. A. Welsh. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In *Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif*, pages 61–110. Springer-Verlag, New York, 1965.

[22] R. v. d. Hofstad. Random graphs and complex networks. lecture notes.

[23] R. v. d. Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, 12:no. 25, 703–766, 2007.

[24] S. Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8:347–361, 7 1999.

[25] S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.

[26] B. Kharlamov. The numbers of generations in a branching process with an arbitrary set of particle types. *Theory of Probability & Its Applications*, 14(3):432–449, 1969.

[27] I. Kolossváry and J. Komjáthy. First passage percolation on inhomogeneous random graphs. *Advances of Applied Probability*, 47, 2015.

[28] C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678–5682, May 2000.

[29] M. Newman and D. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(46):341 – 346, 1999.

[30] M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84:3201–3204, Apr 2000.

[31] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Phys. Rev. E*, 60:7332–7342, Dec 1999.

[32] S. I. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

[33] M. Torok, A. Nelson, L. Alexander, G. C. Mejia, and P. D. MacDonald. Epidemic curves ahead. *Focus on Field Epidemiology*, 1, 2004.

[34] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.